Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (sites)
AUTHORS Kawabata,A., Hikiji,T., Kobatake,N., Inagaki,H., Ikema,Y., Okamoto,S., Okitani,R., Ota,T., Suzuki,Y., Obayashi,M., Nishi,T., Shibahara,T., Tanaka,T., Nakamura,Y., Isogai,T. and Sugano,S.
TITLE NEDO human cDNA sequencing project
JOURNAL Unpublished (2000)
REFERENCE 2 (bases 1 to 2486)
AUTHORS Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T., Shibahara,T., Tanaka,T. and Nakamura,Y.
TITLE Direct Submission
JOURNAL Submitted (29-AUG-2000) Sumio Sugano, Institute of Medical Science, University of Tokyo, Laboratory of Genome Structure Analysis, Human Genome Center; Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan (E-mail:cdna1@ims.u-tokyo.ac.jp, Tel:81-3-5449-5286, Fax:81-3-5449-5416)
COMMENT NEDO human cDNA sequencing project supported by Ministry of International Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology; cDNA library construction, 5'- & 3'-end one pass sequencing: Departent of Virology and Human Genome Center, Institute of Medical Science, University of Tokyo (partly supported by Science and Technology Agency).
FEATURES Location/Qualifiers
source 1. .2486
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="COL05748"
/tissue_type="colon"
/clone_lib="COL"
/note="cloning vector pME18SFL3"
CDS 183. .2282
/note="unnamed protein product"
/codon_start=1
/protein_id="BAB15080.1"
/db_xref="GI:10437626"
/translation="MTSLWREILLESLLGCVSWSLYHDLGPMIYYFPLQTLELTGLEG
FSIAFLSPIFLTITPFWKLVNKKWMLTLLRIITIGSIASFQAPNAKLRLMVLALGVSS
SLIVQAVTWWSGSHLQRYLRIWGFILGQIVLVVLRIWYTSLNPIWSYQMSNKVILTLS
AIATLDRIGTDGDCSKPEEKKTGEVATGMASRPNWLLAGAAFGSLVFLTHWVFGEVSL
VSRWAVSGHPHPGPDPNPFGGAVLLCLASGLMLPSCLWFRGTGLIWWVTGTASAAGLL
YLHTWAAAVSGCVFAIFTASMWPQTLGHLINSGTNPGKTMTIAMIFYLLEIFFCAWCT
AFKFVPGGVYARERSDVLLGTMMLIIGLNMLFGPKKNLDLLLQTKNSSKVLFRKSEKY
MKLFLWLLVGVGLLGLGLRHKAYERKLGKVAPTKEVSAAIWPFRFGYDNEGWSSLERS
AHLLNETGADFITILESDASKPYMGNNDLTMWLGEKLGFYTDFGPSTRYHTWGIMALS
RYPIVKSEHHLLPSPEGEIAPAITLTVNISGKLVDFVVTHFGNHEDDLDRKLQAIAVS
KLLKSSSNQVIFLGYITSAPGSRDYLQLTEHGNVKDIDSTDHDRWCEYIMYRGLIRLG
YARISHAELSDSEIQMAKFRIPDDPTNYRDNQKVVIDHREVSEKIHFNPRFGSYKEGH
NYENNHNFHMNTPKYFL"
BASE COUNT 692 a 516 c 606 g 672 t
ORIGIN

Alignment Scores:

Pred. No.:	1.36e-294	Length:	2486
Score:	3713.00	Matches:	698
Percent Similarity:	100.00%	Conservative:	1
Best Local Similarity:	99.86%	Mismatches:	0
Query Match:	99.81%	Indels:	0
DB:	9	Gaps:	0

US-09-809-638-2 (1-699) x AK025164 (1-2486)

```
Qy      1 MetThrSerLeuTrpArgGluIleLeuLeuGluSerLeuLeuGlyCysValSerTrpSer 20
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    183 ATGACCTCGCTGTGGAGAGAAATCCTCTTGGAGTCGCTGCTGGGATGTGTTTCTTGGTCT 242

Qy     21 LeuTyrHisAspLeuGlyProMetIleTyrTyrPheProLeuGlnThrLeuGluLeuThr 40
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    243 CTCTACCATGACCTGGGACCGATGATCTATTACTTTCCTTTGCAAACACTAGAACTCACT 302

Qy     41 GlyLeuGluGlyPheSerIleAlaPheLeuSerProIlePheLeuThrIleThrProPhe 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    303 GGGCTTGAAGGTTTTAGTATAGCATTTCTTTCTCCAATATTCCTAACAATTACTCCTTTC 362

Qy     61 TrpLysLeuValAsnLysLysTrpMetLeuThrLeuLeuArgIleIleThrIleGlySer 80
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    363 TGGAAATTGGTTAACAAGAAGTGGATGCTAACCCTGCTGAGGATAATCACTATTGGCAGC 422

Qy     81 IleAlaSerPheGlnAlaProAsnAlaLysLeuArgLeuMetValLeuAlaLeuGlyVal 100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    423 ATAGCCTCCTTCCAGGCTCCAAATGCCAAACTTCGACTGATGGTTCTTGCGCTTGGGGTG 482

Qy    101 SerSerSerLeuIleValGlnAlaValThrTrpTrpSerGlySerHisLeuGlnArgTyr 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    483 TCTTCCTCACTGATAGTGCAAGCTGTGACTTGGTGGTCAGGAAGTCATTTGCAAAGGTAC 542

Qy    121 LeuArgIleTrpGlyPheIleLeuGlyGlnIleValLeuValValLeuArgIleTrpTyr 140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    543 CTCAGAATTTGGGGATTCATTTTAGGACAGATTGTTCTTGTTGTTCTACGCATATGGTAT 602

Qy    141 ThrSerLeuAsnProIleTrpSerTyrGlnMetSerAsnLysValIleLeuThrLeuSer 160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    603 ACTTCACTAAACCCAATCTGGAGTTATCAGATGTCCAACAAAGTGATACTGACATTAAGT 662

Qy    161 AlaIleAlaThrLeuAspArgIleGlyThrAspGlyAspCysSerLysProGluGluLys 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    663 GCCATAGCCACACTTGATCGTATTGGCACAGATGGTGACTGCAGTAAACCTGAAGAAAAG 722

Qy    181 LysThrGlyGluValAlaThrGlyMetAlaSerArgProAsnTrpLeuLeuAlaGlyAla 200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    723 AAGACTGGTGAGGTAGCCACGGGGATGGCCTCTAGACCCAACTGGCTGCTGGCAGGGGCT 782

Qy    201 AlaPheGlySerLeuValPheLeuThrHisTrpValPheGlyGluValSerLeuValSer 220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    783 GCTTTTGGTAGCCTTGTGTTCCTCACCCACTGGGTTTTTGGAGAAGTCTCTCTTGTTTCC 842

Qy    221 ArgTrpAlaValSerGlyHisProHisProGlyProAspProAsnProPheGlyGlyAla 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    843 AGATGGGCAGTGAGTGGGCATCCACATCCAGGGCCAGATCCTAACCCATTTGGAGGTGCA 902

Qy    241 ValLeuLeuCysLeuAlaSerGlyLeuMetLeuProSerCysLeuTrpPheArgGlyThr 260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    903 GTACTGCTGTGCTTGGCAAGTGGATTGATGCTTCCATCTTGTTTGTGGTTTCGTGGTACT 962

Qy    261 GlyLeuIleTrpTrpValThrGlyThrAlaSerAlaAlaGlyLeuLeuTyrLeuHisThr 280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    963 GGTTTGATCTGGTGGGTTACAGGAACAGCTTCAGCTGCGGGGCTCCTTTACCTGCACACA 1022

Qy    281 TrpAlaAlaAlaValSerGlyCysValPheAlaIlePheThrAlaSerMetTrpProGln 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1023 TGGGCAGCTGCTGTGTCTGGCTGTGTCTTCGCCATCTTTACTGCATCCATGTGGCCCCAA 1082

Qy    301 ThrLeuGlyHisLeuIleAsnSerGlyThrAsnProGlyLysThrMetThrIleAlaMet 320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1083 ACACTTGGACACCTTATTAACTCAGGGACAAACCCTGGGAAAACCATGACCATTGCCATG 1142

Qy    321 IlePheTyrLeuLeuGluIlePhePheCysAlaTrpCysThrAlaPheLysPheValPro 340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1143 ATATTTTATCTTCTAGAAATATTTTTCTGTGCCTGGTGCACAGCTTTTAAGTTTGTCCCA 1202

Qy    341 GlyGlyValTyrAlaArgGluArgSerAspValLeuLeuGlyThrMetMetLeuIleIle 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1203 GGAGGTGTCTACGCTAGAGAAAGATCAGATGTGCTTTTGGGGACAATGATGTTAATTATC 1262

Qy    361 GlyLeuAsnMetLeuPheGlyProLysLysAsnLeuAspLeuLeuLeuGlnThrLysAsn 380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1263 GGGCTGAATATGCTATTTGGTCCTAAGAAAAACCTTGATTTGCTTCTTCAAACAAAAAAC 1322

Qy    381 SerSerLysValLeuPheArgLysSerGluLysTyrMetLysLeuPheLeuTrpLeuLeu 400
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1323 AGTTCTAAAGTGCTTTTCAGAAAGAGTGAAAAATACATGAAACTTTTTCTGTGGCTGCTT 1382

Qy    401 ValGlyValGlyLeuGlyLeuGlyLeuArgHisLysAlaTyrGluArgLysLeuGly 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1383 GTTGGTGTGGGATTGTTGGGATTAGGACTACGGCATAAAGCCTATGAGAGAAAACTGGGC 1442
```

```
Qy    421 LysValAlaProThrLysGluValSerAlaAlaIleTrpProPheArgPheGlyTyrAsp 440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1443 AAAGTGGCACCAACCAAAGAGGTCTCTGCTGCCATCTGGCCTTTCAGGTTTGGATATGAC 1502

Qy    441 AsnGluGlyTrpSerSerLeuGluArgSerAlaHisLeuLeuAsnGluThrGlyAlaAsp 460
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1503 AATGAAGGGTGGTCTAGTCTAGAAAGATCAGCTCACCTGCTCAATGAAACAGGTGCAGAT 1562

Qy    461 PheIleThrIleLeuGluSerAspAlaSerLysProTyrMetGlyAsnAsnAspLeuThr 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1563 TTCATAACAATTTTGGAGAGTGATGCTTCTAAGCCCTATATGGGGAACAATGACTTAACC 1622

Qy    481 MetTrpLeuGlyGluLysLeuGlyPheTyrThrAspPheGlyProSerThrArgTyrHis 500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1623 ATGTGGCTAGGGGAAAAGTTGGGTTTCTATACAGACTTTGGTCCAAGCACAAGGTATCAC 1682

Qy    501 ThrTrpGlyIleMetAlaLeuSerArgTyrProIleValLysSerGluHisHisLeuLeu 520
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1683 ACTTGGGGGATTATGGCTTTGTCAAGATACCCAATTGTGAAATCTGAGCATCACCTTCTT 1742

Qy    521 ProSerProGluGlyGluIleAlaProAlaIleThrLeuThrValAsnIleSerGlyLys 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1743 CCGTCACCAGAGGGCGAGATCGCACCAGCCATCACATTGACCGTTAACATTTCGGGCAAG 1802

Qy    541 LeuValAspPheValValThrHisPheGlyAsnHisGluAspAspLeuAspArgLysLeu 560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1803 CTGGTGGATTTTGTCGTGACACACTTTGGGAACCACGAAGATGACCTCGACAGGAAACTG 1862

Qy    561 GlnAlaIleAlaValSerLysLeuLeuLysSerSerSerAsnGlnValIlePheLeuGly 580
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1863 CAGGCTATTGCTGTTTCAAAACTACTGAAAAGTAGCTCTAATCAAGTGATATTTCTGGGA 1922

Qy    581 TyrIleThrSerAlaProGlySerArgAspTyrLeuGlnLeuThrGluHisGlyAsnVal 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1923 TATATCACTTCAGCACCTGGCTCCAGAGATTATCTACAGCTCACTGAACATGGCAATGTG 1982

Qy    601 LysAspIleAspSerThrAspHisAspArgTrpCysGluTyrIleMetTyrArgGlyLeu 620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1983 AAGGATATCGACAGCACTGATCATGACAGATGGTGTGAATACATTATGTATCGAGGGCTG 2042

Qy    621 IleArgLeuGlyTyrAlaArgIleSerHisAlaGluLeuSerAspSerGluIleGlnMet 640
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2043 ATCAGGTTGGGTTATGCAAGAATCTCCCATGCTGAACTGAGTGATTCAGAAATTCAGATG 2102

Qy    641 AlaLysPheArgIleProAspAspProThrAsnTyrArgAspAsnGlnLysValValIle 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2103 GCAAAATTTAGGATCCCTGATGACCCCACTAATTATAGAGACAACCAGAAAGTGGTCATA 2162

Qy    661 AspHisArgGluValSerGluLysIleHisPheAsnProArgPheGlySerTyrLysGlu 680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2163 GACCACAGAGAAGTTTCTGAGAAAATTCATTTTAATCCCAGATTTGGATCCTACAAAGAA 2222

Qy    681 GlyHisAsnTyrGluAsnAsnHisHisPheHisMetAsnThrProLysTyrPheLeu 699
          ||||||||||||||||||||||||:::||||||||||||||||||||||||||||||
Db   2223 GGACACAATTATGAAAACAACCATAATTTTCATATGAATACTCCCAAATACTTTTTA 2279

RESULT 2
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 21, 2002, 22:11:54 ; Search time 3431 Seconds
(without alignments)
4263.378 Million cell updates/sec

Title: US-09-809-638-2
Perfect score: 3720
Sequence: 1 MTSLWREILLESLLGCVSWS.....EGHNYENNHHFHMNTPKYFL 699

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=x1h
-Q=/cgn2_1/USPTO_spool/US09809638/runat_21102002_153504_1831/app_query.fasta_1.839
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09809638_@CGN_1_1_2231_@runat_21102002_153504_1831 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htgo_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	3713	99.8	2486	9	AK025164	AK025164 Homo sapi
2	3125	84.0	2412	10	BC016523	BC016523 Mus muscu
c 3	760	20.4	316613	8	SCCHRIII	X59720 S.cerevisia
4	639	17.2	34034	8	SPAC688	AL355632 S.pombe c
5	625.5	16.8	1234	8	D89130	D89130 Schizosacch
6	495	13.3	995	11	CNS06JWV	AL402149 T3 end of
7	471	12.7	184206	9	AC020593	AC020593 Homo sapi
8	287.5	7.7	222885	2	AL158816	AL158816 Homo sapi
c 9	268.5	7.2	222885	2	AL158816	AL158816 Homo sapi
10	264.5	7.1	18342	8	SPAC589	AL512496 S.pombe c
c 11	182.5	4.9	165893	2	AC106576	AC106576 Rattus no
c 12	182	4.9	305	6	AX337478	AX337478 Sequence
13	152.5	4.1	849	11	CNS06EJ2	AL395172 T3 end of
c 14	144.5	3.9	225532	2	AC096364	AC096364 Rattus no
c 15	141.5	3.8	31717	1	SCJ21	AL109747 Streptomy
c 16	139	3.7	1348	9	HUMTPIPSA	K03223 Human TPI (
17	138	3.7	17013	1	EAAMASL	X77921 E.amylovara
c 18	134	3.6	9720	1	PSDNGC	Z73914 Pseudomonas
19	133	3.6	274050	1	AL627276	AL627276 Salmonell
20	132	3.5	29593	8	AF402142	AF402142 Hyaloraph
21	131.5	3.5	7330	1	AB026907	AB026907 Ochrobact
c 22	130.5	3.5	240050	1	AL627267	AL627267 Salmonell
23	130	3.5	193050	1	AJ414149	AJ414149 Yersinia
24	129	3.5	96532	9	CNS07EG6	AL590079 Human chr
c 25	128	3.4	879	6	A85835	A85835 Sequence 49
c 26	128	3.4	879	6	AR155328	AR155328 Sequence
c 27	128	3.4	879	6	E65853	E65853 Genome DNA
c 28	128	3.4	11299	1	AE005976	AE005976 Caulobact
29	128	3.4	20396	1	AE008827	AE008827 Salmonell
c 30	128	3.4	133863	9	AL157709	AL157709 Human DNA
31	127.5	3.4	206045	2	AL669826	AL669826 Mus muscu
32	127	3.4	4715	1	AF039919	AF039919 Brevibact
33	126	3.4	1340	9	S73482	S73482 uterine wat
c 34	126	3.4	26720	1	AE008724	AE008724 Salmonell
35	125.5	3.4	101648	1	AF339846	AF339846 Microscil
36	125	3.4	1340	6	AR002457	AR002457 Sequence
37	125	3.4	1340	6	AR028509	AR028509 Sequence
38	125	3.4	1340	6	AX331552	AX331552 Sequence
39	125	3.4	1340	6	AX331945	AX331945 Sequence
40	125	3.4	1340	9	HUMCHIP28A	M77829 Human chann
41	125	3.4	1657	9	HSU41517	U41517 Human chann
42	125	3.4	2764	9	BC022486	BC022486 Homo sapi
c 43	125	3.4	30132	1	SC10G8	AL158057 Streptomy
44	124.5	3.3	1647	6	AX120510	AX120510 Sequence
45	124.5	3.3	1770	6	AX066343	AX066343 Sequence

ALIGNMENTS

RESULT 1
AK025164
LOCUS AK025164 2486 bp mRNA linear PRI 29-SEP-2000
DEFINITION Homo sapiens cDNA: FLJ21511 fis, clone COL05748.
ACCESSION AK025164
VERSION AK025164.1 GI:10437625
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Homo sapiens colon cDNA to mRNA, clone_lib:COL clone:COL05748.
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (sites)
AUTHORS Kawabata,A., Hikiji,T., Kobatake,N., Inagaki,H., Ikema,Y., Okamoto,S., Okitani,R., Ota,T., Suzuki,Y., Obayashi,M., Nishi,T., Shibahara,T., Tanaka,T., Nakamura,Y., Isogai,T. and Sugano,S.
TITLE NEDO human cDNA sequencing project
JOURNAL Unpublished (2000)
REFERENCE 2 (bases 1 to 2486)
AUTHORS Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T., Shibahara,T., Tanaka,T. and Nakamura,Y.
TITLE Direct Submission
JOURNAL Submitted (29-AUG-2000) Sumio Sugano, Institute of Medical Science, University of Tokyo, Laboratory of Genome Structure Analysis, Human Genome Center; Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan (E-mail:cdna1@ims.u-tokyo.ac.jp, Tel:81-3-5449-5286, Fax:81-3-5449-5416)
COMMENT NEDO human cDNA sequencing project supported by Ministry of International Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology; cDNA library construction, 5'- & 3'-end one pass sequencing: Departent of Virology and Human Genome Center, Institute of Medical Science, University of Tokyo (partly supported by Science and Technology Agency).
FEATURES Location/Qualifiers
source 1..2486
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="COL05748"
/tissue_type="colon"
/clone_lib="COL"
/note="cloning vector pME18SFL3"
CDS 183..2282
/note="unnamed protein product"
/codon_start=1
/protein_id="BAB15080.1"
/db_xref="GI:10437626"
/translation="MTSLWREILLESLLGCVSWSLYHDLGPMIYYFPLQTLELTGLEG
FSIAFLSPIFLTITPFWKLVNKKWMLTLLRIITIGSIASFQAPNAKLRLMVLALGVSS
SLIVQAVTWWSGSHLQRYLRIWGFILGQIVLVVLRIWYTSLNPIWSYQMSNKVILTLS
AIATLDRIGTDGDCSKPEEKKTGEVATGMASRPNWLLAGAAFGSLVFLTHWVFGEVSL
VSRWAVSGHPHPGPDPNPFGGAVLLCLASGLMLPSCLWFRGTGLIWWVTGTASAAGLL
YLHTWAAAVSGCVFAIFTASMWPQTLGHLINSGTNPGKTMTIAMIFYLLEIFFCAWCT
AFKFVPGGVYARERSDVLLGTMMLIIGLNMLFGPKKNLDLLLQTKNSSKVLFRKSEKY
MKLFLWLLVGVGLLGLGLRHKAYERKLGKVAPTKEVSAAIWPFRFGYDNEGWSSLERS
AHLLNETGADFITILESDASKPYMGNNDLTMWLGEKLGFYTDFGPSTRYHTWGIMALS
RYPIVKSEHHLLPSPEGEIAPAITLTVNISGKLVDFVVTHFGNHEDDLDRKLQAIAVS
KLLKSSSNQVIFLGYITSAPGSRDYLQLTEHGNVKDIDSTDHDRWCEYIMYRGLIRLG
YARISHAELSDSEIQMAKFRIPDDPTNYRDNQKVVIDHREVSEKIHFNPRFGSYKEGH
NYENNHNFHMNTPKYFL"
BASE COUNT 692 a 516 c 606 g 672 t
ORIGIN

Alignment Scores:

Pred. No.:	1.36e-294	Length:	2486
Score:	3713.00	Matches:	698
Percent Similarity:	100.00%	Conservative:	1
Best Local Similarity:	99.86%	Mismatches:	0
Query Match:	99.81%	Indels:	0
DB:	9	Gaps:	0

US-09-809-638-2 (1-699) x AK025164 (1-2486)

```
Qy      1 MetThrSerLeuTrpArgGluIleLeuLeuGluSerLeuLeuGlyCysValSerTrpSer 20
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    183 ATGACCTCGCTGTGGAGAGAAATCCTCTTGGAGTCGCTGCTGGGATGTGTTTCTTGGTCT 242

Qy     21 LeuTyrHisAspLeuGlyProMetIleTyrTyrPheProLeuGlnThrLeuGluLeuThr 40
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    243 CTCTACCATGACCTGGGACCGATGATCTATTACTTTCCTTTGCAAACACTAGAACTCACT 302

Qy     41 GlyLeuGluGlyPheSerIleAlaPheLeuSerProIlePheLeuThrIleThrProPhe 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    303 GGGCTTGAAGGTTTTAGTATAGCATTTCTTTCTCCAATATTCCTAACAATTACTCCTTTC 362

Qy     61 TrpLysLeuValAsnLysLysTrpMetLeuThrLeuLeuArgIleIleThrIleGlySer 80
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    363 TGGAAATTGGTTAACAAGAAGTGGATGCTAACCCTGCTGAGGATAATCACTATTGGCAGC 422

Qy     81 IleAlaSerPheGlnAlaProAsnAlaLysLeuArgLeuMetValLeuAlaLeuGlyVal 100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    423 ATAGCCTCCTTCCAGGCTCCAAATGCCAAACTTCGACTGATGGTTCTTGCGCTTGGGGTG 482

Qy    101 SerSerSerLeuIleValGlnAlaValThrTrpTrpSerGlySerHisLeuGlnArgTyr 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    483 TCTTCCTCACTGATAGTGCAAGCTGTGACTTGGTGGTCAGGAAGTCATTTGCAAAGGTAC 542

Qy    121 LeuArgIleTrpGlyPheIleLeuGlyGlnIleValLeuValValLeuArgIleTrpTyr 140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    543 CTCAGAATTTGGGGATTCATTTTAGGACAGATTGTTCTTGTTGTTCTACGCATATGGTAT 602

Qy    141 ThrSerLeuAsnProIleTrpSerTyrGlnMetSerAsnLysValIleLeuThrLeuSer 160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    603 ACTTCACTAAACCCAATCTGGAGTTATCAGATGTCCAACAAAGTGATACTGACATTAAGT 662

Qy    161 AlaIleAlaThrLeuAspArgIleGlyThrAspGlyAspCysSerLysProGluGluLys 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    663 GCCATAGCCACACTTGATCGTATTGGCACAGATGGTGACTGCAGTAAACCTGAAGAAAAG 722

Qy    181 LysThrGlyGluValAlaThrGlyMetAlaSerArgProAsnTrpLeuLeuAlaGlyAla 200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    723 AAGACTGGTGAGGTAGCCACGGGGATGGCCTCTAGACCCAACTGGCTGCTGGCAGGGGCT 782

Qy    201 AlaPheGlySerLeuValPheLeuThrHisTrpValPheGlyGluValSerLeuValSer 220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    783 GCTTTTGGTAGCCTTGTGTTCCTCACCCACTGGGTTTTTGGAGAAGTCTCTCTTGTTTCC 842

Qy    221 ArgTrpAlaValSerGlyHisProHisProGlyProAspProAsnProPheGlyGlyAla 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    843 AGATGGGCAGTGAGTGGGCATCCACATCCAGGGCCAGATCCTAACCCATTTGGAGGTGCA 902

Qy    241 ValLeuLeuCysLeuAlaSerGlyLeuMetLeuProSerCysLeuTrpPheArgGlyThr 260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    903 GTACTGCTGTGCTTGGCAAGTGGATTGATGCTTCCATCTTGTTTGTGGTTTCGTGGTACT 962

Qy    261 GlyLeuIleTrpTrpValThrGlyThrAlaSerAlaAlaGlyLeuLeuTyrLeuHisThr 280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    963 GGTTTGATCTGGTGGGTTACAGGAACAGCTTCAGCTGCGGGGCTCCTTTACCTGCACACA 1022

Qy    281 TrpAlaAlaAlaValSerGlyCysValPheAlaIlePheThrAlaSerMetTrpProGln 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1023 TGGGCAGCTGCTGTGTCTGGCTGTGTCTTCGCCATCTTTACTGCATCCATGTGGCCCCAA 1082

Qy    301 ThrLeuGlyHisLeuIleAsnSerGlyThrAsnProGlyLysThrMetThrIleAlaMet 320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1083 ACACTTGGACACCTTATTAACTCAGGGACAAACCCTGGGAAAACCATGACCATTGCCATG 1142

Qy    321 IlePheTyrLeuLeuGluIlePhePheCysAlaTrpCysThrAlaPheLysPheValPro 340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1143 ATATTTTATCTTCTAGAAATATTTTTCTGTGCCTGGTGCACAGCTTTTAAGTTTGTCCCA 1202

Qy    341 GlyGlyValTyrAlaArgGluArgSerAspValLeuLeuGlyThrMetMetLeuIleIle 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1203 GGAGGTGTCTACGCTAGAGAAAGATCAGATGTGCTTTTGGGGACAATGATGTTAATTATC 1262

Qy    361 GlyLeuAsnMetLeuPheGlyProLysLysAsnLeuAspLeuLeuLeuGlnThrLysAsn 380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1263 GGGCTGAATATGCTATTTGGTCCTAAGAAAAACCTTGATTTGCTTCTTCAAACAAAAAAC 1322

Qy    381 SerSerLysValLeuPheArgLysSerGluLysTyrMetLysLeuPheLeuTrpLeuLeu 400
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1323 AGTTCTAAAGTGCTTTTCAGAAAGAGTGAAAAATACATGAAACTTTTTCTGTGGCTGCTT 1382

Qy    401 ValGlyValGlyLeuLeuGlyLeuGlyLeuArgHisLysAlaTyrGluArgLysLeuGly 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1383 GTTGGTGTGGGATTGTTGGGATTAGGACTACGGCATAAAGCCTATGAGAGAAAACTGGGC 1442
```

```
Qy     421 LysValAlaProThrLysGluValSerAlaAlaIleTrpProPheArgPheGlyTyrAsp 440
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1443 AAAGTGGCACCAACCAAAGAGGTCTCTGCTGCCATCTGGCCTTTCAGGTTTGGATATGAC 1502

Qy     441 AsnGluGlyTrpSerSerLeuGluArgSerAlaHisLeuLeuAsnGluThrGlyAlaAsp 460
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1503 AATGAAGGGTGGTCTAGTCTAGAAAGATCAGCTCACCTGCTCAATGAAACAGGTGCAGAT 1562

Qy     461 PheIleThrIleLeuGluSerAspAlaSerLysProTyrMetGlyAsnAsnAspLeuThr 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1563 TTCATAACAATTTTGGAGAGTGATGCTTCTAAGCCCTATATGGGGAACAATGACTTAACC 1622

Qy     481 MetTrpLeuGlyGluLysLeuGlyPheTyrThrAspPheGlyProSerThrArgTyrHis 500
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1623 ATGTGGCTAGGGGAAAAGTTGGGTTTCTATACAGACTTTGGTCCAAGCACAAGGTATCAC 1682

Qy     501 ThrTrpGlyIleMetAlaLeuSerArgTyrProIleValLysSerGluHisHisLeuLeu 520
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1683 ACTTGGGGGATTATGGCTTTGTCAAGATACCCAATTGTGAAATCTGAGCATCACCTTCTT 1742

Qy     521 ProSerProGluGlyGluIleAlaProAlaIleThrLeuThrValAsnIleSerGlyLys 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1743 CCGTCACCAGAGGGCGAGATCGCACCAGCCATCACATTGACCGTTAACATTTCGGGCAAG 1802

Qy     541 LeuValAspPheValValThrHisPheGlyAsnHisGluAspAspLeuAspArgLysLeu 560
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1803 CTGGTGGATTTTGTCGTGACACACTTTGGGAACCACGAAGATGACCTCGACAGGAAACTG 1862

Qy     561 GlnAlaIleAlaValSerLysLeuLeuLysSerSerSerAsnGlnValIlePheLeuGly 580
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1863 CAGGCTATTGCTGTTTCAAAACTACTGAAAAGTAGCTCTAATCAAGTGATATTTCTGGGA 1922

Qy     581 TyrIleThrSerAlaProGlySerArgAspTyrLeuGlnLeuThrGluHisGlyAsnVal 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1923 TATATCACTTCAGCACCTGGCTCCAGAGATTATCTACAGCTCACTGAACATGGCAATGTG 1982

Qy     601 LysAspIleAspSerThrAspHisAspArgTrpCysGluTyrIleMetTyrArgGlyLeu 620
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1983 AAGGATATCGACAGCACTGATCATGACAGATGGTGTGAATACATTATGTATCGAGGGCTG 2042

Qy     621 IleArgLeuGlyTyrAlaArgIleSerHisAlaGluLeuSerAspSerGluIleGlnMet 640
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2043 ATCAGGTTGGGTTATGCAAGAATCTCCCATGCTGAACTGAGTGATTCAGAAATTCAGATG 2102

Qy     641 AlaLysPheArgIleProAspAspProThrAsnTyrArgAspAsnGlnLysValValIle 660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2103 GCAAAATTTAGGATCCCTGATGACCCCACTAATTATAGAGACAACCAGAAAGTGGTCATA 2162

Qy     661 AspHisArgGluValSerGluLysIleHisPheAsnProArgPheGlySerTyrLysGlu 680
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2163 GACCACAGAGAAGTTTCTGAGAAAATTCATTTTAATCCCAGATTTGGATCCTACAAAGAA 2222

Qy     681 GlyHisAsnTyrGluAsnAsnHisHisPheHisMetAsnThrProLysTyrPheLeu 699
           |||||||||||||||||||||||:::||||||||||||||||||||||||||||||
Db    2223 GGACACAATTATGAAAACAACCATAATTTTCATATGAATACTCCCAAATACTTTTTA 2279
```

RESULT 2
BC016523

```
LOCUS       BC016523               2412 bp    mRNA    linear   ROD 05-NOV-2001
DEFINITION  Mus musculus, Similar to hypothetical protein FLJ21511, clone
            MGC:27925 IMAGE:3584006, mRNA, complete cds.
ACCESSION   BC016523
VERSION     BC016523.1  GI:16741399
KEYWORDS    MGC.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 2412)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-OCT-2001) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Gilbert Smith, Ph.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Baylor College of Medicine Human Genome
            Sequencing Center
            Center code: BCM-HGSC
            Web site: http://www.hgsc.bcm.tmc.edu/cdna/
            Contact: villalon@bcm.tmc.edu.
            Villalon, D.K., Luna, R.A., Hale, S.M., Hulyk, S., Lu, X., Garcia,
            A.M., Holloway, M., Telford, B, Hodgson, A., Bouck, J., Yu, W.,
            Muzny,D.M., Gibbs,R.A.

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 35 Row: k Column: 5
            This clone was selected for full length sequencing because it
            passed the following selection criteria: Similarity but not
            identity to protein.
FEATURES              Location/Qualifiers
     source           1. .2412
                      /organism="Mus musculus"
                      /db_xref="taxon:10090"
                      /clone="MGC:27925 IMAGE:3584006"
                      /tissue_type="Mammary tumor. Metallothionien-TGF alpha
                      model. 10 month old virgin mouse. Taken by biopsy."
                      /clone_lib="NCI_CGAP_Mam1"
                      /lab_host="DH10B"
                      /note="Vector: pCMV-SPORT6"
     CDS              113. .2212
                      /codon_start=1
                      /product="Similar to hypothetical protein FLJ21511"
                      /protein_id="AAH16523.1"
                      /db_xref="GI:16741400"
                      /translation="MPGLWRAIALETLLGYVSWSLYHGLSPMIYYFPLQTLELTGLEF
                      FCVAFLSPILLTIPPLWKLVNKKWTLSLLRIVTVGSIASFEAPNAKLRLMVLALGVSS
                      SLIVQTVTWWSGSGLQRYLKIWGFILGHVLLLVLRIWYTSLNPIWSYQMSNRVILTLS
                      AVAVLDRIGTDGDYRNPEGKKPREVATGRTSLSSWLLPGAAFGSLLFLTHWIFGEVSI
                      VSRWAVSGHPHPGPDPNPFGGAVLLGFSSGLMLSGSSWLHDAGLAWWMTGAASAMGLL
                      YLRTWAAAVSGCVLAVFTGSMWPQVLGHLVNSGKNSGEAMATGMILYVLQTFFCAWCT
                      AFKFVPGGVYARERSDVLLGTIMVIIGLSMLFGPKKNLDFLLQTKNSPKTLLRCSEKY
                      MKLILWLFVGVGLLGLGLRHRTYERQLGRGAPATVVSAAIWPFRFGYDNEGWPNLERS
                      AQLLKETGADFITILESDASKPYIGNNDLTMWLGEKLGFYTDFGPSTRDHTWGIMVLS
                      RYPIVRSEHHLLPSPEGEIAPAITMTVNVSNRLVDFVVTHFGNHEDDLDRKLQAIAVS
                      KLLKNCSNQVIFLGYITSEPGSRDYIQLTKHGNVKDIDSSDGDRWCEYIMYRGLIRLG
                      YARISHAELSDSEIQMAKFRIPDDPANYRDNQKVVIDHRGVPKNIHFNPRFGSYKEGH
                      NYENTHHFHMNTPKYFV"
BASE COUNT      628 a    561 c    616 g    607 t
ORIGIN
```

Alignment Scores:

Pred. No.:	1.9e-246	Length:	2412
Score:	3125.00	Matches:	579
Percent Similarity:	88.70%	Conservative:	41
Best Local Similarity:	82.83%	Mismatches:	79
Query Match:	84.01%	Indels:	0
DB:	10	Gaps:	0

US-09-809-638-2 (1-699) x BC016523 (1-2412)

```
Qy       1 MetThrSerLeuTrpArgGluIleLeuLeuGluSerLeuLeuGlyCysValSerTrpSer 20
           |||      |||||||||   |||  ||||||:::|||||||||   ||||||||||||
Db     113 ATGCCAGGCCTGTGGAGAGCAATTGCCCTGGAGACCTTGCTGGGCTACGTCTCCTGGTCT 172

Qy      21 LeuTyrHisAspLeuGlyProMetIleTyrTyrProProLeuGlnThrLeuGluLeuThr 40
           |||||||||   |||   ||||||||||||||||||||||||||||||||||||||||
Db     173 CTCTACCACGGACTGAGCCCCATGATCTATTATTTTCCCCTGCAAACGCTGGAGCTCACA 232
```

```
Qy      41 GlyLeuGluGlyPheSerIleAlaPheLeuSerProIlePheLeuThrIleThrProPhe 60
Db     233 GGGCTGGAATTCTTTTGTGTAGCCTTTCTCTCTCCGATACTCTTAACAATTCCCCCACTC 292

Qy      61 TrpLysLeuValAsnLysLysTrpMetLeuThrLeuLeuArgIleIleThrIleGlySer 80
Db     293 TGGAAACTCGTCAATAAGAAGTGGACACTCTCTCTGCTGAGGATCGTCACTGTAGGCAGC 352

Qy      81 IleAlaSerPheGlnAlaProAsnAlaLysLeuArgLeuMetValLeuAlaLeuGlyVal 100
Db     353 ATAGCTTCCTTTGAGGCTCCAAATGCCAAGCTTCGACTGATGGTCCTCGCCCTTGGTGTG 412

Qy     101 SerSerSerLeuIleValGlnAlaValThrTrpTrpSerGlySerHisLeuGlnArgTyr 120
Db     413 TCTTCTTCCTTGATAGTGCAAACTGTGACCTGGTGGTCAGGAAGTGGATTGCAAAGGTAC 472

Qy     121 LeuArgIleTrpGlyPheIleLeuGlyGlnIleValLeuValValLeuArgIleTrpTyr 140
Db     473 CTCAAAATCTGGGGATTCATTCTGGGACACGTCCTTCTTCTTGTCCTCCGAATATGGTAC 532

Qy     141 ThrSerLeuAsnProIleTrpSerTyrGlnMetSerAsnLysValIleLeuThrLeuSer 160
Db     533 ACATCATTAAACCCAATTTGGAGCTATCAGATGTCCAACAGAGTGATTCTGACATTGAGT 592

Qy     161 AlaIleAlaThrLeuAspArgIleGlyThrAspGlyAspCysSerLysProGluGluLys 180
Db     593 GCCGTAGCTGTGCTTGACCGAATTGGCACAGATGGTGACTACAGAAACCCTGAGGGGAAG 652

Qy     181 LysThrGlyGluValAlaThrGlyMetAlaSerArgProAsnTrpLeuLeuAlaGlyAla 200
Db     653 AAGCCCCGTGAGGTTGCCACGGGCAGGACCTCTCTATCTAGTTGGCTGTTGCCAGGCGCC 712

Qy     201 AlaPheGlySerLeuValPheLeuThrHisTrpValPheGlyGluValSerLeuValSer 220
Db     713 GCCTTTGGGAGCCTCCTGTTCCTAACTCACTGGATCTTCGGAGAGGTCTCCATTGTTTCC 772

Qy     221 ArgTrpAlaValSerGlyHisProHisProGlyProAspProAsnProPheGlyGlyAla 240
Db     773 AGATGGGCAGTCAGTGGTCATCCCCACCCAGGGCCAGATCCTAATCCTTTTGGAGGTGCG 832

Qy     241 ValLeuLeuCysLeuAlaSerGlyLeuMetLeuProSerCysLeuTrpPheArgGlyThr 260
Db     833 GTTCTACTGGGCTTCTCAAGTGGACTGATGCTATCGGGTTCATCATGGCTTCATGATGCT 892

Qy     261 GlyLeuIleTrpTrpValThrGlyThrAlaSerAlaAlaGlyLeuLeuTyrLeuHisThr 280
Db     893 GGGTTAGCCTGGTGGATGACAGGAGCAGCTTCGGCCATGGGTCTCCTCTACCTGCGCACA 952

Qy     281 TrpAlaAlaAlaValSerGlyCysValPheAlaIlePheThrAlaSerMetTrpProGln 300
Db     953 TGGGCAGCTGCTGTTTCTGGATGTGTCCTGGCAGTCTTTACGGGGTCTATGTGGCCTCAA 1012

Qy     301 ThrLeuGlyHisLeuIleAsnSerGlyThrAsnProGlyLysThrMetThrIleAlaMet 320
Db    1013 GTACTTGGCCATCTTGTGAACTCAGGGAAAAACTCCGGAGAAGCCATGGCCACTGGCATG 1072

Qy     321 IlePheTyrLeuLeuGluIlePhePheCysAlaTrpCysThrAlaPheLysPheValPro 340
Db    1073 ATCCTTTATGTTTTACAAACATTTTTCTGTGCCTGGTGCACAGCGTTTAAGTTTGTTCCC 1132

Qy     341 GlyGlyValTyrAlaArgGluArgSerAspValLeuLeuGlyThrMetMetLeuIleIle 360
Db    1133 GGAGGTGTCTACGCTAGAGAAAGATCTGATGTGCTTTTGGGGACCATAATGGTAATTATT 1192

Qy     361 GlyLeuAsnMetLeuPheGlyProLysLysAsnLeuAspLeuLeuLeuGlnThrLysAsn 380
Db    1193 GGATTGAGTATGCTGTTTGGACCTAAGAAAAACCTTGACTTTCTTCTTCAAACAAAAAAC 1252

Qy     381 SerSerLysValLeuPheArgLysSerGluLysTyrMetLysLeuPheLeuTrpLeuLeu 400
Db    1253 AGTCCTAAAACGCTTTTGAGATGCAGTGAAAAATACATGAAACTCATTCTGTGGCTGTTT 1312

Qy     401 ValGlyValGlyLeuLeuGlyLeuGlyLeuArgHisLysAlaTyrGluArgLysLeuGly 420
Db    1313 GTTGGTGTGGGGTTACTGGGATTAGGGCTACGGCACAGAACCTACGAGAGGCAGTTAGGC 1372

Qy     421 LysValAlaProThrLysGluValSerAlaAlaIleTrpProPheArgPheGlyTyrAsp 440
Db    1373 CGAGGGGCACCAGCGACAGTGGTCTCGGCTGCCATCTGGCCTTTTAGGTTTGGATATGAC 1432

Qy     441 AsnGluGlyTrpSerSerLeuGluArgSerAlaHisLeuLeuAsnGluThrGlyAlaAsp 460
Db    1433 AATGAAGGCTGGCCCAATCTAGAGAGGTCTGCTCAACTGCTCAAGGAGACAGGTGCAGAT 1492

Qy     461 PheIleThrIleLeuGluSerAspAlaSerLysProTyrMetGlyAsnAsnAspLeuThr 480
Db    1493 TTTATCACCATTTTGGAGAGCGATGCTTCCAAGCCCTATATAGGGAACAACGACTTAACA 1552

Qy     481 MetTrpLeuGlyGluLysLeuGlyPheTyrThrAspPheGlyProSerThrArgTyrHis 500
Db    1553 ATGTGGCTGGGGGAGAAGCTGGGTTTCTACACAGACTTTGGGCCAAGCACCAGGGATCAC 1612

Qy     501 ThrTrpGlyIleMetAlaLeuSerArgTyrProIleValLysSerGluHisHisLeuLeu 520
Db    1613 ACCTGGGGGATTATGGTGCTGTCTCGGTACCCGATTGTCAGATCGGAACATCACCTTCTT 1672

Qy     521 ProSerProGluGlyGluIleAlaProAlaIleThrLeuThrValAsnIleSerGlyLys 540
Db    1673 CCGTCGCCAGAGGGCGAGATCGCACCAGCCATAACCATGACAGTTAACGTCTCCAACAGA 1732

Qy     541 LeuValAspPheValValThrHisPheGlyAsnHisGluAspAspLeuAspArgLysLeu 560
Db    1733 CTGGTGGATTTTGTGGTGACACACTTTGGGAATCATGAAGATGACCTTGACAGGAAGCTA 1792

Qy     561 GlnAlaIleAlaValSerLysLeuLeuLysSerSerSerAsnGlnValIlePheLeuGly 580
Db    1793 CAGGCTATTGCAGTTTCAAAACTGCTGAAAAATTGTTCAAATCAAGTGATATTTCTGGGT 1852

Qy     581 TyrIleThrSerAlaProGlySerArgAspTyrLeuGlnLeuThrGluHisGlyAsnVal 600
Db    1853 TATATCACTTCAGAGCCTGGTTCCAGAGATTACATACAACTCACAAAACATGGCAACGTG 1912

Qy     601 LysAspIleAspSerThrAspHisAspArgTrpCysGluTyrIleMetTyrArgGlyLeu 620
Db    1913 AAGGATATTGACAGCTCAGATGGAGACAGATGGTGTGAGTACATCATGTACCGGGGTCTG 1972

Qy     621 IleArgLeuGlyTyrAlaArgIleSerHisAlaGluLeuSerAspSerGluIleGlnMet 640
Db    1973 ATCAGGTTGGGCTATGCAAGAATCTCTCATGCAGAACTGAGTGACTCTGAGATTCAGATG 2032

Qy     641 AlaLysPheArgIleProAspAspProThrAsnTyrArgAspAsnGlnLysValValIle 660
Db    2033 GCCAAATTCAGGATCCCGGATGATCCCGCCAATTACAGAGACAACCAGAAAGTCGTCATA 2092

Qy     661 AspHisArgGluValSerGluLysIleHisPheAsnProArgPheGlySerTyrLysGlu 680
Db    2093 GACCACCGAGGAGTTCCCAAGAACATTCACTTCAACCCCAGATTTGGCTCCTACAAAGAA 2152

Qy     681 GlyHisAsnTyrGluAsnAsnHisHisPheHisMetAsnThrProLysTyrPheLeu 699
Db    2153 GGACACAATTATGAGAACACCCATCATTTTCACATGAATACACCCAAATACTTCGTT 2209

RESULT 3
SCCHRIII/c
LOCUS       SCCHRIII              316613 bp    DNA     linear   PLN 29-JUN-2001
DEFINITION  S.cerevisiae chromosome III complete DNA sequence.
ACCESSION   X59720 S43845 S49180 S58084 S93798
VERSION     X59720.2  GI:14588895
KEYWORDS    chromosome.
SOURCE      baker's yeast.
  ORGANISM  Saccharomyces cerevisiae
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE   1  (bases 1 to 316613)
  AUTHORS   Rad,M.R., Lutzenkirchen,K., Xu,G., Kleinhans,U. and Hollenberg,C.P.
  TITLE     The complete sequence of a 11,953 bp fragment from C1G on
```

chromosome III encompasses four new open reading frames
JOURNAL Yeast 7 (5), 533-538 (1991)
MEDLINE 91377317
REFERENCE 2 (bases 1 to 316613)
AUTHORS Wilson,C., Bergantino,E., Lanfranchi,G., Valle,G., Carignani,G. and Frontali,L.
TITLE A putative serine/threonine protein kinase gene on chromosome III of Saccharomyces cerevisiae
JOURNAL Yeast 8 (1), 71-77 (1992)
MEDLINE 92254506
REFERENCE 3 (bases 1 to 316613)
AUTHORS Biteau,N., Fremaux,C., Hebrard,S., Menara,A., Aigle,M. and Crouzet,M.
TITLE The complete sequence of a 10.8kb fragment to the right of the chromosome III centromere of Saccharomyces cerevisiae
JOURNAL Yeast 8 (1), 61-70 (1992)
MEDLINE 92254505
REFERENCE 4 (bases 1 to 316613)
AUTHORS Benit,P., Chanet,R., Fabre,F., Faye,G., Fukuhara,H. and Sor,F.
TITLE Sequence of the sup61-RAD18 region on chromosome III of Saccharomyces cerevisiae
JOURNAL Yeast 8 (2), 147-153 (1992)
MEDLINE 92221691
REFERENCE 5 (bases 1 to 316613)
AUTHORS Bolle,P.A., Gilliquet,V., Berben,G., Dumont,J. and Hilger,F.
TITLE The complete sequence of K3B, a 7.9 kb fragment between PGK1 and CRY1 on chromosome III, reveals the presence of seven open reading frames
JOURNAL Yeast 8 (3), 205-213 (1992)
MEDLINE 92245758
REFERENCE 6 (bases 1 to 316613)
AUTHORS Sor,F., Cheret,G., Fabre,F., Faye,G. and Fukuhara,H.
TITLE Sequence of the HMR region on chromosome III of Saccharomyces cerevisiae
JOURNAL Yeast 8 (3), 215-222 (1992)
MEDLINE 92245759
REFERENCE 7 (bases 1 to 316613)
AUTHORS Skala,J., Purnelle,B. and Goffeau,A.
TITLE The complete sequence of a 10.8 kb segment distal of SUF2 on the right arm of chromosome III from Saccharomyces cerevisiae reveals seven open reading frames including the RVS161, ADP1 and PGK genes
JOURNAL Yeast 8 (5), 409-417 (1992)
MEDLINE 92327849
REFERENCE 8 (bases 1 to 316613)
AUTHORS Oliver,S.G., van der Aart,Q.J., Agostoni-Carbone,M.L., Aigle,M., Alberghina,L., Alexandraki,D., Antoine,G., Anwar,R., Ballesta,J.P., Benit,P. et al.
TITLE The complete DNA sequence of yeast chromosome III
JOURNAL Nature 357 (6373), 38-46 (1992)
MEDLINE 92244356
REFERENCE 9 (bases 1 to 316613)
AUTHORS Scherens,B., Messenguy,F., Gigot,D. and Dubois,E.
TITLE The complete sequence of a 9,543 bp segment on the left arm of chromosome III reveals five open reading frames including glucokinase and the protein disulfide isomerase
JOURNAL Yeast 8 (7), 577-585 (1992)
MEDLINE 92397595
REFERENCE 10 (bases 1 to 316613)
AUTHORS Wilson,C., Grisanti,P. and Frontali,L.
TITLE The complete sequence of a 6146 bp fragment of Saccharomyces cerevisiae chromosome III contains two new open reading frames
JOURNAL Yeast 8 (7), 569-575 (1992)
MEDLINE 92397594
REFERENCE 11 (bases 26740 to 32076)
AUTHORS Defoor,E., Debrabandere,R., Keyers,B., Voet,M. and Volckaert,G.
TITLE Nucleotide sequence of D10B, a BamHI fragment on the small-ring chromosome III of Saccharomyces cerevisiae
JOURNAL Yeast 8 (8), 681-687 (1992)
MEDLINE 93070606
REFERENCE 12 (bases 169581 to 171116; 171683 to 172169)
AUTHORS Agostoni Carbone,M.L., Panzeri,L., Muzi Falconi,M., Carcano,C., Plevani,P. and Lucchini,G.
TITLE Nucleotide sequence of 9.2 kb left of CRY1 on yeast chromosome III from strain AB972: evidence for a Ty insertion and functional analysis of open reading frame YCR28
JOURNAL Yeast 8 (9), 805-812 (1992)
MEDLINE 93070619
REFERENCE 13 (bases 1 to 315338)
AUTHORS Valle,G.
TITLE TA-repeat microsatellites are closely associated with ARS consensus sequences in yeast chromosome III
JOURNAL Yeast 9 (7), 753-759 (1993)
MEDLINE 93377412
REFERENCE 14 (bases 1 to 315338)
AUTHORS Slonimski,P.P. and Brouillet,S.
TITLE A data-base of chromosome III of Saccharomyces cerevisiae
JOURNAL Yeast 9 (9), 941-1029 (1993)
MEDLINE 94091061
REFERENCE 15 (bases 1 to 315339)
AUTHORS Rodriguez-Cousino,N., Lill,R., Neupert,W. and Court,D.A.
TITLE Identification and initial characterization of the cytosolic protein Ycr77p
JOURNAL Yeast 11 (6), 581-585 (1995)
MEDLINE 95373282
REFERENCE 16 (bases 1 to 316613)
AUTHORS MIPS.
TITLE Direct Submission
JOURNAL Submitted (16-MAR-1992) MIPS, D-8033 Martinsried, FRG. Data collected by MIPS on behalf of the European Yeast Chromosome III Sequencing project
REMARK replaced by [13]
REFERENCE 17 (bases 1 to 314957)
AUTHORS Jimenez,A.
TITLE Direct Submission
JOURNAL Submitted (28-DEC-1992) MIPS, D-8033 Martinsried, FRG. Data collected by MIPS on behalf of the European Yeast Chromosome III Sequencing project. Update originating from A. Jimenez
REMARK replaced by [14]
REFERENCE 18 (bases 1 to 316613)
AUTHORS Louis,E.J.
TITLE Direct Submission
JOURNAL Submitted (25-JUN-1993) MIPS, D-8033 Martinsried, FRG. Data collected by MIPS on behalf of the European Yeast Chromosome III Sequencing project. Update originating from E.J. Louis
REMARK revised by [18]
REFERENCE 19 (bases 1 to 316613)
AUTHORS Louis,E.J.
TITLE Direct Submission
JOURNAL Submitted (14-FEB-1995) MIPS, D-8033 Martinsried, FRG. Data collected by MIPS on behalf of the European Yeast Chromosome III Sequencing project. Update originating from E.J. Louis
REMARK Revised by [20]
REFERENCE 20 (bases 1 to 316613)
AUTHORS Gromadka,R.
TITLE Direct Submission
JOURNAL Submitted (29-JAN-1996) R. Gromadka, Protein Biosynthesis, Institute of Biochemistry and Biophysics, Pawinskiego 5A, Warsaw, Poland, Electronic Mail Address: robert@psd.ibb.waw.pl
REMARK revised by [21]
REFERENCE 21 (bases 1 to 316613)
AUTHORS MIPS.
TITLE Direct Submission
JOURNAL Submitted (17-JUN-2001) MIPS Yeast Genome Database, GSF - Ingolstaedter Landstrasse 1 D-85764 Neuherberg, Germany, mips-yeast-adm@gsf.de
REMARK Resequencing project, achieved by the joint effort of G. Valles and G. Volckaerts laboratories. Munich information center for protein sequences.
COMMENT On Jul 3, 2001 this sequence version replaced gi:1907116.
The contig sequence which has been released to the file server in 1992 has been subject to a resequencing project, achieved by the joint effort of G. Valles and G. Volckaerts laboratories. The following lines present a summary of the altered entities. The resequenced chrIII contig is 316613 bp with GCG-check: 429 GENETIC ENTITIES valid no longer:
YCL012w (YCL014w + YCL012w -> YCL014w / C_B1367 + C_C231->C_A1636),

PEP Y, ORF in identical form now C-terminal part of YCL014w. (A frame shift in
the nt sequence of YCL014w leads to an elongated polypeptide.) This elongated polypeptide encoded by the current YCL014w resembles the Bud3p polypeptide from the
EMBL-Entry (Accession:SC175801;ID:U17580, total length 1636 aa)
YCL006c, PEP Y, ORF obsolete, is not an ORF in current Chromosome III contig.
YCR062w (YCR061w + YCR062w -> YCR061w / C_A583 + C_B120 -> C_B631), PEP Y, ORF in identical form now C-terminal part of YCR061w.
YCR068w-a (YCR068w + YCR068w-a -> YCR068w / C_C429 + C_A145 -> C_A520), PEP Y, ORF in identical form now C-terminal part of YCR068w.
YCR103c (C_F111, former ORF in region corresponding to coord. 307255-307587(C)) is obsolete, open reading frame is destroyed by single nt exchange resulting in a
stop codon after 45 aa.
NEW GENETIC ENTITIES:
YCL26c-b (C_E193(YCL027c-a)), PEP Y, new ORF
YCL021w-a (C_C125), PEP Y, new orf
ALTERED GENETIC ENTITIES:
YCL076w, PEP Y, peptide elongated N-terminally by 51 aa YCL074w, PEP Y, 3 aa exchanged
YCL073c, PEP Y, 4 aa exchanged
YCL068c (C_F190->C_E260), PEP Y, peptide elongated N-terminally by 70 aa YCL065w, PEP Y, 1 aa exchanged
YCL064c, PEP Y, 3 aa exchanged
YCL061c (C_D853->C_F1096), PEP Y, peptide elongated N-terminally by 243 aa YCL051w, PEP Y, 11 aa altered, 6 aa exchanged, 1 aa additional, 4 aa less at the C-terminus
YCL050c, PEP Y, 1 aa exchanged
YCL049c, PEP Y, 4 aa exchanged
YCL042w, PEP Y, 8 aa altered, 7 aa exchanged, 1 aa additional
YCL034w (C_A265->C_B354), PEP Y, peptide elongated N-terminally by 89 aa YCL030c, PEP Y, 1 aa exchanged
YCL028w, PEP Y, 1 aa exchanged
YCL027w, PEP Y, 3 aa exchanged
YCL026c-a, PEP Y, 8 aa altered, 7 aa exchanged, 1 aa additional
YCL025c (C_F633->C_F595), PEP Y, 7 aa exchanged, the c-terminus is shorter by 38 aa and the last 5 aa are altered
LTR deltaCL05, 1 nt exchange
LTR deltaCL04, 1 nt deletion
TY2A_C, PEP Y, 1 aa exchanged

Alignment Scores:

Pred. No.:	6.24e-50	Length:	316613
Score:	760.00	Matches:	228
Percent Similarity:	47.20%	Conservative:	126
Best Local Similarity:	30.40%	Mismatches:	256
Query Match:	20.43%	Indels:	140
DB:	8	Gaps:	31

[redacted]-809-638-2 (1-699) x SCCHRIII (1-316613)

```
Qy      15 GlyCysValSerTrpSerLeuTyrHisAspLeuGlyProMetIleTyrTyrPheProLeu 34
           |||   :::   ||||||           |||         |||::::::||||||||
Db 146794 GGTTTTCTTTTCTGGTCGAACGTTACGTCCCTGTTATGTAGTATTTGGCATTTCCCGCTA 146735

Qy      35 GlnThrLeuGluLeuThrGlyLeuGluGlyPheSerIleAlaPheLeuSerProIlePhe 54
              :::   ::::::|||   |||         :::   :::|||   |||||||||
Db 146734 TGGTATATGGGAATCTCAGGTTATGAAGCTGCAATATTGGGTTATTTGGGACCCATTTTC 146675

Qy      55 LeuThrIleThrProPheTrpLysLeuValAsnLysLysTrpMetLeuThrLeuLeuArg 74
           |||   :::   ||||||                 ::::::   :::   |||
Db 146674 TTATATCTG---CCGTTCGTTTCTGAAGCCTTCACGCAATATGGTGTACTTTTAGGAGGT 146618

Qy      75 IleIleThrIleGlySerIleAlaSerPheGlnAlaProAsnAlaLysLeuArgLeuMet 94
           ||||||   ||||||:::           |||   |||       :::||||||||:::
Db 146617 ATTATTGCCATTGGTGCCTATATT---GTTCAGATGCCA------GAATTAAGGTTGATT 146567

Qy      95 ValLeuAlaLeuGlyValSerSerSerLeu--------------------------Ile 105
              :::|||:::|||   |||   ::::::                          |||
Db 146566 TCTGTAGCTGTGGGAACTTCCATTACCGTTGCAACGTTTGTACAAAATCTAAGATATATC 146507

Qy     106 ValGlnAlaValThrTrpTrpSerGlySerHisLeuGlnArgTyrLeuArgIleTrpGly 125
           |||   |||   :::|||                 :::         |||
Db 146506 ACAAATGCGGAGACTAGTTTCTCT----------------TTTGCTCTAACTTGG--- 146468

Qy     126 PheIleLeuGlyGlnIleValLeuValValLeuArgIleTrpTyrThrSerLeuAsnPro 145
              :::||||||   :::      |||:::|||::::::   :::   :::   ||||||
Db 146467 ---CTGTTAGGTCTTGTTGCATCTGTGATCTTGAAAATGGGGTTCTATACCAACAATCCA 146411

Qy     146 IleTrp----------------------SerTyrGlnMetSerAsnLysValIleLeu 157
           |||                        |||         ||||||      |||
Db 146410 ACTTGGGTCATTTTAGATGAACGTAATGGTGGGTAT---------AATAAGACAGCTCTC 146360

Qy     158 ThrLeuSerAlaIle--------------AlaThrLeuAspArgIleGlyThrAspGly 172
           |||:::   :::                 ::::::   |||      :::|||
Db 146359 GTGCTTACTGTTTTATTCGGCATGCTGTCGCCTTATGTTAATTCAATTAATTTCGAAGGG 146300

Qy     173 AspCysSerLysProGluGluLysLysThrGlyGluValAlaThrGlyMetAlaSerArg 192
                           |||:::      :::      :::      ||||||
Db 146299 --------------------AAAAGGAATGCTCAAGCAAAATCT------GCTTCGTTG 146267

Qy     193 ProAsnTrpLeuLeuAlaGlyAlaAlaPheGlySerLeuValPheLeuThrHisTrpVal 212
              |||                |||||||||||:::|||      |||   :::
Db 146266 ATTGGCAAATTATTTTTGGCTGTTGGTTTTGGCTCGTTGTTATTCGGAATTCATCAGTTA 146207

Qy     213 PheGlyGluValSerLeuValSerArgTrpAlaValSerGlyHisProHisPro---Gly 231
              :::   |||            ||||||      |||:::            |||
Db 146206 TTGACGGATTCTTCTACTACTATTTATTGGGCATGGGAAGGTTACAATGAATCACACGGT 146147

Qy     232 ProAspProAsnProPheGlyGly------AlaValLeuLeuCysLeuAlaSer------ 247
           |||   |||   |||:::|||                ::::::      ||||||
Db 146146 CCCTTGCCATGGCCTTGGGGCGCCTTAACTTGTACGGTCATGTTATTTGCTTCTTTGAGT 146087

Qy     248 ---------------GlyLeuMetLeuProSerCysLeuTrpPheArgGlyThrGlyLeu 262
                         |||      |||      ||||||                |||
Db 146086 TCTGTGAAGTTTATGGGCAAGCCATTAGTTCCATGTTTG-----------------TTG 146045

Qy     263 IleTrpTrpValThrGlyThrAlaSerAla------------------------------ 272
           :::         |||         ||||||
Db 146044 CTTCTCATATCCACTGCTGTACTTTCAGCTAGAAGCATTACACAATGGCCTAAATATATT 145985

Qy     273 ---AlaGlyLeuLeuTyrLeuHisThrTrpAlaAlaAlaValSerGlyCysValPheAla 291
              ||||||||||||            |||   |||:::         |||   :::
Db 145984 TTCGGTGGTTTATTGTAC----------GCTATCGCTATGCTTTGGTTAGTTCCTTCG 145937

Qy     292 IlePheThrAla--------------SerMetTrpProGlnThrLeuGlyHisLeuIle 306
           |||:::|||          ::::::|||         |||
Db 145936 TATTTTTCTGCATTAGGCCAAGTTCAAAACATATGGGTTTATGTCCTA----------- 145889

Qy     307 AsnSerGlyThrAsnProGlyLysThrMetThrIleAlaMetIlePheTyrLeuLeuGlu 326
                                   :::   ::::::   :::||||||   |||
Db 145888 ------------------------TCATTCTCCGTTTATATTATCTTTGTCCTTGCCCAT 145853

Qy     327 IlePhePheCysAlaTrpCysThrAlaPheLysPheValProGlyGlyValTyrAlaArg 346
           :::            |||      |||:::   |||||||||   |||         |||
Db 145852 GTT-----------TGGGTCGTTGCATACGCATTTGTTCCAATGGGCTGGGTACTGAGG 145805

Qy     347 GluArgSerAspValLeuLeu------GlyThrMetMetLeuIleIleGlyLeuAsnMet 364
           |||:::   :::   :::|||         :::      :::||||||||||
Db 145804 GAGAAGATTGAGACGGTTCTTGCCTTTTCTTCCACATTTATCATTATTGGTGCTTTAACA 145745

Qy     365 LeuPheGlyProLysLysAsnLeuAspLeuLeuLeuGlnThrLysAsnSerSerLysVal 384
                                |||||||||:::::   |||   |||
Db 145744 TGC-----------AAAAACCTTAACATTCAACTGGTGACTATGGGCAAA--------- 145706

Qy     385 LeuPheArgLysSerGluLysTyrMetLysLeuPheLeuTrpLeu---LeuValGlyVal 403
                                        |||   |||:::::::::         |||
Db 145705 --------------------------AAATTCTTCATTTATGTTTTCTTCTTTGCCGTG 145673

Qy     404 GlyLeuLeuGlyLeuGlyLeuArgHisLysAlaTyrGluArgLysLeuGlyLysValAla 423
           ||||||   |||      |||         |||      |||
Db 145672 GCCCTGCTATCACTAACAGCTAGGTTCGTGTATGATATTAGACCTACAGGAATTCCTCAG 145613

Qy     424 Pro------------ThrLysGluValSerAlaAlaIleTrpProPheArgPheGlyTyr 439
```

```
Db 145612 CCTTATCATCCAGATTCTCAGTTGATTACAGCTGGTATTTGGACTATCCACTTTGGTCTC 145553

Qy    440 AspAsnGluGlyTrpSerSerLeuGluArgSerAlaHisLeuLeuAsnGluThrGlyAla 459
Db 145552 GATAATGATATGTGGGCATCTGAAGACAGAATGATCAACCTTATTAAAGATATGGAACTA 145493

Qy    460 AspPheIleThrIleLeuGluSerAspAlaSerLysProTyrMetGlyAsnAsnAspLeu 479
Db 145492 GATGTGGTAGGTCTACTAGAAACAGATACACAAAGAATTACCATGGGGAACAGGGATCTA 145433

Qy    480 ThrMetTrpLeuGlyGluLysLeuGlyPheTyrThrAspPheGlyProSerThrArgTyr 499
Db 145432 ACTAGCAAACTAGCTCATGATTTGAATATGTATGCAGATTTCGGACCAGGTCCAAATAAA 145373

Qy    500 HisThrTrpGlyIleMetAlaLeuSerArgTyrProIleValLysSerGluHisHisLeu 519
Db 145372 CATACCTGGGGCTGTGTTCTTCTTTCTAAATTCCCTATCGTAAATTCTACGCATCATTTA 145313

Qy    520 LeuProSerProGluGlyGluIleAlaProAlaIleThrLeuThrVal---AsnIleSer 538
Db 145312 TTGCCCTCTCCAGTTGGGGAACTTGCGCCAGCCATTCATGCCACACTTCAAACGTACAAT 145253

Qy    539 GlyLysLeuValAspPheValValThrHisPheGlyAsnHisGluAspAspLeuAspArg 558
Db 145252 GACACTCTCGTTGACGTCTTTGTATTCCATAGTGGACAAGAAGAGGATGAAGAGGATAGA 145193

Qy    559 LysLeuGlnAlaIleAlaValSerLysLeuLeuLysSerSerSerAsnGlnValIlePhe 578
Db 145192 AGACTGCAAAGTAACTACATGGCTAAGCTCATGGGCAATACGACTCGCCCAGCTATTTTA 145133

Qy    579 LeuGlyTyrIleThrSerAlaProGlySerArgAspTyr---LeuGlnLeuThrGluHis 597
Db 145132 TTAAGTTACTTAGTTGTTGATCCAGGTGAAGGCAACTACAATACGTACGTTAGTGAAACA 145073

Qy    598 GlyAsnValLysAspIleAspSerThrAspHisAspArgTrpCysGluTyrIleMetTyr 617
Db 145072 TCCGGAATGCACGACATTGATCCTTCTGACGATGATAGATGGTGTGAGTATATCTTGTAT 145013

Qy    618 ArgGlyLeuIleArgLeuGlyTyrAlaArgIleSerHisAlaGluLeuSerAspSerGlu 637
Db 145012 AGGGGCTTGAGAAGAACAGGATATGCTAGAGTTGCAAGAGGAACGATAACCGATACGGAG 144953

Qy    638 IleGlnMetAlaLysPheArgIleProAspAspProThrAsnTyrArgAspAsnGlnLys 657
Db 144952 CTACAAGTTGGTAAGTTCCAAGTTTTG----------AGTGAGCAA 144917

Qy    658 ValValIleAspHisArgGlu----------ValSerGluLysIleHis 670
Db 144916 GCGTTAGTAGAGCACTCGGATTCTATGTATGAATACGGTCATATGAGTGAACCGGAATAT 144857

Qy    671 PheAsnProArgPheGlySerTyrLysGluGlyHisAsnTyrGluAsnAsnHisHisPhe 690
Db 144856 GAGGACATGAAATTTCCAGATAAGTTTTTAGGCGAAGGT---GAGAGGGGTCACTTCTAC 144800

Qy    691 HisMet---AsnThrProLysTyrPheLeu 699
Db 144799 CATGTTTTTGATGAGCCACGTTATTACTTA 144770
```

RESULT 4
SPAC688

```
LOCUS       SPAC688                34034 bp    DNA     linear   PLN 08-MAY-2000
DEFINITION  S.pombe chromosome I cosmid c688.
ACCESSION   AL355632
VERSION     AL355632.1  GI:7768476
KEYWORDS    Actin-related protein; cell polarization; coiled-coil; cut9
            interacting protein; dna polymerase zeta catalytic subunit; drug
            sensitivity; DUF51; endocytosis internalization phase; ENTH domain;
            glutathione S transferase; I/LWEQ domain; Kornberg's mediator (SRB)
            subcomplex; membrane cytoskeleton assembly; microtubule stability;
            mitochondrial carrier protein; Protein of unknown function domain;
            RNA polymerase II holoenzyme; scn1; Tf1-type LTR; yeast SLA2; yeast
            SRB8; yeast YCR017C.
SOURCE      fission yeast.
  ORGANISM  Schizosaccharomyces pombe
            Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
            Schizosaccharomycetales; Schizosaccharomycetaceae;
            Schizosaccharomyces.
REFERENCE   1  (bases 1 to 34034)
  AUTHORS   Aert,R., Robben,J., Volckaert,G., Wood,V., Rajandream,M.A. and
            Barrell,B.G.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-APR-2000) European Schizosaccharomyces genome
            sequencing project, Sanger Centre, The Wellcome Trust Genome
            Campus, Hinxton, Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk
            and Katholieke Universiteit Leuven, Laboratory of Gene Technology,
            Kardinaal Mercierlaan 92, B-3001 Leuven, Belgium
COMMENT     Notes:
            Details of yeast sequencing at the Sanger Centre are available on
            the World Wide Web.
            (URL, http://www.sanger.ac.uk/Projects/S_pombe/)
            During 1995 to 1996 about 66% of S. pombe chromosome 1 was
            sequenced by the Sanger Centre.  The sequencing of the S. pombe
            genome is now being continued with funding from The European
            Commission. Fourteen European sequencing laboratories, including
            the Sanger Centre, are participating in the project.
            Protein coding regions (CDS) have been predicted with the help of
            computer analysis using the Genefinder program in PomBase (an ACEDB
            database) with additional predictions for the branch-acceptor sites
            supplied by the program Sp3splice. CAUTION: It is possible that for
            any individual CDS we may have underestimated or overestimated the
            number of introns/exons or we may not have chosen the correct
            splice donor/acceptor sites.
            CDS are numbered using the following system eg SPBC25H2.01c. SP (S.
            pombe), B (chromosome 2), c25H2 (cosmid name), .01 (first CDS), c
            (complementary strand).
            The more significant matches with motifs in the PROSITE database
            are also included but some of these may be fortuitous.
            The length in codons is given for each CDS.
            IMPORTANT: This sequence MAY NOT be the entire insert of the
            sequenced clone.  It may be shorter because we only sequence
            overlapping sections once, or longer, because we arrange for a
            small overlap between neighbouring submissions.
            Cosmid c688 is overlapped at the 3' end by cosmid c3G9, EMBL entry
            SPAC3G9, accession number AL021046.
FEATURES              Location/Qualifiers
     source           1. .34034
                      /organism="Schizosaccharomyces pombe"
                      /strain="972h-"
                      /db_xref="taxon:4896"
                      /chromosome="I"
                      /map="IL"
                      /clone="cosmid c688"
     gene             1. .1280
                      /gene="SPAC688.01"
     CDS              1. .1280
                      /partial
                      /gene="SPAC688.01"
                      /note="SPAC688.01, len:425, SIMILARITY:Saccharomyces
                      cerevisiae, YCQ7_YEAST, hypothetical 107.9 kd protein.,
                      (953 aa), fasta scores: opt: 1034, E():0, (46.5% identity
                      in 430 aa)"
                      /codon_start=3
                      /label=SPAC688.01
                      /product="hypothetical protein similar to yeast YCR017C
                      protein involved in sensitivity to certain drugs"
                      /protein_id="CAB90768.1"
                      /db_xref="GI:7768477"
                      /translation="SYFMYINHGWCSYLGGLIFTSYVLIYSFASIRISSFYSPAKVWG
                      GAFLVYILYSLAHVWVVAYEFVPGGPILRERTSYILIFIGWNLAALVPAYSGESKEPN
                      KADSSVVDIKQSDSSYRRRSFKKSLLTGFCLALMALKFAIQNMPPYDYTPYHPNEKLF
                      TAGIWTIHFGLDNFMYASENRIRDAVRDMELDVFGLLESDTQRLIMGFRDLTQVLAHD
                      LGMYADYGPGPDKHTWGAALLSKFPIVNSTHHLLPSPQGELAPAIHATLDVYGELIDV
                      VVSHNGQYESQLDRRLQSTELARIMRESPRPLVFLGYVVSNVGQEPQTILTRDTGMLD
                      IEPADYDRWCQYIFYRGVKRIGYARLHRSTITDTELQTGKFLVTKDLGRNVRIDKEHV
                      PESHRYPSLFEGTGVNGHYYDNNLVVHEPWYYD"
```

```
gene            complement(join(1545. .1979,2109. .2256,2343. .2392))
                /gene="SPAC688.02c"
CDS             complement(join(1545. .1979,2109. .2256,2343. .2392))
                /gene="SPAC688.02c"
                /note="SPAC688.02c, len:210, SIMILARITY:Saccharomyces
                cerevisiae, Q12143, chromosome xvi reading frame orf
                ypl233w., (216 aa), fasta scores: opt: 183, E():9.1e-05,
                (22.7% identity in 185 aa)"
                /codon_start=1
                /label=SPAC688.02c
                /product="hypothetical protein with very low similarity to
                yeast ypl233w"
                /protein_id="CAB90769.1"
                /db_xref="GI:7768478"
                /translation="MSSHGDASVLPKIALESINDYSYVKSVFRQAIQEKLAIHLPEQA
                SREAGMGREDQLRIRVKEMLNELIESTFSIIQENVTINGFDANSALQDPKNSDQIEPF
                DLALRTRVQQLFNEVEDAHVLVARYRKSVPAQYEKAYVDAMEQQTAFLRNVKDDYVSL
                QDKEVENPDEQTSTTVFRKEDIDRYEQTIAKVAYLKKNLPRVVARLEKTP"
misc_feature    complement(1980. .1990)
                /gene="SPAC688.02c"
                /note="ttaacatatag, splice branch and acceptor"
misc_feature    complement(2103. .2108)
                /gene="SPAC688.02c"
                /note="gtgagt, splice donor sequence"
misc_feature    complement(2257. .2271)
                /gene="SPAC688.02c"
                /note="ttaacactttataag, splice branch and acceptor"
misc_feature    complement(2337. .2342)
                /gene="SPAC688.02c"
                /note="gtgtgt, splice donor sequence"
gene            complement(join(2596. .2953,3083. .3297,3381. .3422))
                /gene="SPAC688.03c"
CDS             complement(join(2596. .2953,3083. .3297,3381. .3422))
                /gene="SPAC688.03c"
                /note="SPAC688.03c, len:204, SIMILARITY:Homo sapiens,
                Q9Y4X0, ammecr1 protein., (333 aa), fasta scores: opt:
                443, E():3.4e-25, (40.3% identity in 186 aa)"
                /codon_start=1
                /label=SPAC688.03c
                /product="conserved hypothetical protein"
                /protein_id="CAB90770.1"
                /db_xref="GI:7768479"
                /translation="MICWPSLNKESWNDKEYCYYCFEVVAATLEHRKVRDKWNAKSWT
                RSIPLFVKFASGKGHDKQLRGCIGTFRARPLVTNLTYFSKQAAFCDERFRPISLGELA
                LLECQIDLLVDFEPIDDPLDWEVGIHGVSIKFTANGIRYSSTYLPSVAAEQRWDQEET
                LESLIHKAGYYGSIRSLQITATRYKSLEIGCTYEEYLHNLELLG"
misc_feature    complement(join(2662. .2953,3083. .3273))
                /gene="SPAC688.03c"
                /note="Match to PF01871 DUF51, Protein of unknown function
                Score 49.41"
misc_feature    complement(2954. .2976)
                /gene="SPAC688.03c"
                /note="ctaatacaacattttaaaaatag, splice branch and
                acceptor"
misc_feature    complement(3077. .3082)
                /gene="SPAC688.03c"
                /note="gtaagt, splice donor sequence"
misc_feature    complement(3298. .3315)
                /gene="SPAC688.03c"
                /note="ctaatataaatggacaag, splice branch and acceptor"
misc_feature    complement(3375. .3380)
                /gene="SPAC688.03c"
                /note="gtatga, splice donor sequence"
gene            complement(4047. .4775)
                /gene="SPAC688.04c"
CDS             complement(4047. .4775)
                /gene="SPAC688.04c"
                /note="SPAC688.04c, len:242, SIMILARITY:Saccharomyces
                cerevisiae, YIV8_YEAST, Glutathione transferase GTT1, (234
                aa), fasta scores: opt: 274, E():3.5e-11, (27.0% identity
                in 222 aa)"
                /codon_start=1
                /label=SPAC688.04c
                /product="putative glutathione S transferase"
                /protein_id="CAB90771.1"
                /db_xref="GI:7768480"
                /translation="MIVLHHLKNSRSTRIVWMLEELKVPYEIKVYDRVDGRAPPAYTK
                LSPLGKSPIVVDDGVTYIESAAILEHLVRKYGPSFKPSEEDVAELEKYELWMHFSEAS
                LMPFIWASHVLDLSVNMTPIFFRYIVRQFVNGIKSKYLSKETFLNLDYIDNHLASNEY
                FAGEQFTAADPQMCFPIFAAQRDYLSQKPYKNIKRWMRVVSDRPACRIAAEKVEDNTL
                TLFSDVERYSHPPTPPPEQVRSDE"
misc_feature    complement(4164. .4775)
                /gene="SPAC688.04c"
                /note="Match to PF00043 GST, Glutathione S-transferases.
                Score 85.13"
gene            complement(join(5637. .6842,6927. .6980))
                /gene="SPAC688.06c"
CDS             complement(join(5637. .6842,6927. .6980))
                /gene="SPAC688.06c"
                /note="SPAC688.06c, len:419, predicted to contain
                coiled-coil region, SIMILARITY:S. cerevisiae, LOW to
                Actin-related protein, regulator of microtubule stability,
                YB61_YEAST, (324 aa), fasta scores: opt: 125, E():0.76,
                (21.3% identity in 300 aa)"
                /codon_start=1
                /label=SPAC688.06c
                /product="hypothetical protein very putative Actin-related
                protein, regulator of microtubule stability"
                /protein_id="CAB90772.1"
                /db_xref="GI:7768481"
                /translation="MSAEEYIEVSSSPDIFTDDDDMITIEPELNKNPKDCNSKRKRSV
                TECCEIRLITSKCDFESTQQLVHHNCTGHKVHEHNLNAVDEEDFDTENLPLLFSSFSD
                NESDILEPDLNTRVAEDNDVLLSRYSKIKNSASCRNTFEHSAYHSNREEISSSGFYYH
                RKPQLFEKSLEKLGNKSIEANRSPLIKELCESANSTENVCFSVSTVDEIQQRHPSAGH
                SIDSTCQSNSFLEGDSATHKKKKTDNIKEFTSCEFNDRSRTLLNYAGYMDTNKNADNE
                AKSLKEKLENFPVEKLRAIAESYGFKSSDSKATLIKIVESCLDAIDSRSQSKKLGKET
                PHDYLITSTKTVLEFDDIVTQTHRAISQVVKQAKDNSVWIKILTYSAIDVEEFQLWLK
                RKNLNVSLDLIKSWCDKYGVLMKGSWH"
misc_feature    complement(6843. .6857)
                /gene="SPAC688.06c"
                /note="ttaactgagctttag, splice branch and acceptor"
misc_feature    complement(6921. .6926)
                /gene="SPAC688.06c"
                /note="gtatga, splice donor sequence"
gene            complement(7359. .10475)
```

```
Alignment Scores:
Pred. No.:              2.64e-41        Length:         34034
Score:                  639.00          Matches:        152
Percent Similarity:     50.00%          Conservative:   70
Best Local Similarity:  34.23%          Mismatches:     178
Query Match:            17.18%          Indels:         44
DB:                     8               Gaps:           10

US-09-809-638-2 (1-699) x SPAC688 (1-34034)

Qy    276 LeuTyrLeu---HisThrTrpAlaAlaAlaValSerGlyCysValPheAlaIlePheThr 294
          :::|||:::   |||   |||   :::   :::   |||   :::|||      :::
Db     12 ATGTATATTAACCATGGATGGTGTAGTTATTTGGGTGGGTTAATATTTACTTCCTACGTT 71

Qy    295 --------------AlaSerMetTrpProGlnThrLeuGlyHisLeuIleAsnSerGly 309
                       ||||||:::                        |||:::|||
Db     72 TTAATTTACTCGTTCGCTTCTATC--------------------AGAATTTCCTCCTTC 110

Qy    310 ThrAsnProGlyLysThrMetThrIleAlaMetIlePheTyrLeuLeuGluIlePhePhe 329
             :::|||   |||            |||   :::   |||:::|||
Db    111 TATTCTCCTGCGAAGGTTTGGGGTGGCGCTTTTTTGGTTTACATCCTTTACTCATTAGCT 170

Qy    330 CysAlaTrpCysThrAlaPheLysPheValProGlyGlyValTyrAlaArgGluArgSer 349
               |||      |||:::::|||||||||||||||         |||||||||:::
Db    171 CATGTTTGGGTCGTAGCCTACGAGTTTGTTCCAGGTGGACCCATTCTTCGAGAGAGAACT 230

Qy    350 AspValLeuLeuGlyThrMetMetLeuIleIleGlyLeuAsnMetLeuPhe--------- 366
               :::|||            :::   ||||||   |||:::
Db    231 TCTTACATTCTA------------ATTTTCATTGGTTGGAACCTTGCAGCTTTAGTACCT 278
```

```
Qy    367 --------------------GlyProLysLysAsnLeuAspLeuLeuLeuGlnThrLys 379
                              |||   |||           ::::::      |||
Db    279 GCGTATTCGGGTGAGTCCAAAGAACCCAACAAAGCAGACTCATCTGTAGTAGATATCAAA 338

Qy    380 AsnSerSerLysValLeuPheArgLysSerGluLysTyrMetLysLeuPheLeuTrpLeu 399
          |||               |||:::|||   |||                        |||
Db    339 CAATCAGATTCTTCTTATCGTAGAAGAAGTTTCAAAAAATCA--------------CTA 383

Qy    400 LeuValGlyValGlyLeuLeuGlyLeuGlyLeuArgHisLysAlaTyrGluArgLysLeu 419
          |||   |||      |||      :::   |||:::
Db    384 TTGACAGGTTTTTGTCTGGCTCTTATGGCCCTAAAATTTGCCATTCAGAATATGCCTCCG 443

Qy    420 GlyLysValAlaPro------------ThrLysGluValSerAlaAlaIleTrpProPhe 435
                    |||               |||      :::|||   ||||||
Db    444 TATGACTATACACCTTACCATCCAAATGAAAAGCTTTTCACTGCTGGTATTTGGACTATT 503

Qy    436 ArgPheGlyTyrAspAsnGluGlyTrpSerSerLeuGluArgSerAlaHisLeuLeuAsn 455
          ||||||  ||||||      ::::::|||      |||            :::
Db    504 CACTTTGGCTTAGACAATTTCATGTATGCAAGTGAGAATCGTATACGCGATGCTGTTCGA 563

Qy    456 GluThrGlyAlaAspPheIleThrIleLeuGluSerAspAlaSerLysProTyrMetGly 475
          :::         |||         :::||||||||||||      :::      ||||||
Db    564 GATATGGAACTTGATGTCTTTGGTCTACTGGAATCTGACACACAACGTTTAATTATGGGA 623

Qy    476 AsnAsnAspLeuThrMetTrpLeuGlyGluLysLeuGlyPheTyrThrAspPheGlyPro 495
          ||||||||      |||         ||||||  |||  |||:::||||||
Db    624 TTTCGAGATTTAACCCAGGTTTTAGCACATGATCTAGGAATGTATGCGGATTATGGCCCA 683

Qy    496 SerThrArgTyrHisThrTrpGlyIleMetAlaLeuSerArgTyrProIleValLysSer 515
          ||||||||||||          ||||||::::::|||||||||   |||
Db    684 GGCCCTGATAAACATACTTGGGGCGCTGCGTTACTCTCCAAATTTCCTATTGTCAACTCG 743

Qy    516 GluHisHisLeuLeuProSerProGluGlyGluIleAlaProAlaIleThrLeuThrVal 535
          ||||||||||||||||||||:::||||||:::||||||||||||      |||:::
Db    744 ACTCACCATTTGTTACCTTCTCCTCAAGGAGAACTTGCTCCTGCTATTCATGCAACACTA 803

Qy    536 AsnIleSerGlyLysLeuValAspPheValValThrHisPheGlyAsnHisGluAspAsp 555
          ::::::   |||:::|||:::|||   ||||||:::|||   |||   :::|||
Db    804 GATGTCTACGGAGAGCTAATAGATGTTGTGGTTTCTCACAATGGTCAATATGAAAGTCAG 863

Qy    556 LeuAspArgLysLeuGlnAlaIleAlaValSerLysLeuLeuLysSerSerSerAsnGln 575
          |||||||||:::|||||||:::      ::::::::::::::::::   |||
Db    864 TTGGATCGACGTCTTCAGAGCACTGAGTTGGCTCGAATTATGCGCGAGAGTCCAAGGCCT 923

Qy    576 ValIlePheLeuGlyTyrIleThrSerAlaProGlySerArgAspTyrLeuGlnLeuThr 595
          ::::::|||||||||||||:::   |||      |||                |||||
Db    924 CTAGTGTTTCTCGGATATGTTGTTTCAAACGTTGGCCAAGAACCCCAAACTATTCTAACG 983

Qy    596 GluHisGlyAsnValLysAspIleAspSerThrAspHisAspArgTrpCysGluTyrIle 615
                   :::   ||||||:::      |||:::||||||||||||:::||||||
Db    984 CGAGACACGGGAATGTTAGACATTGAACCTGCTGACTATGATCGTTGGTGTCAGTATATA 1043

Qy    616 MetTyrArgGlyLeuIleArgLeuGlyTyrAlaArgIleSerHisAlaGluLeuSerAsp 635
          |||||||||:::   |||:::||||||||||||:::      :::   ::::::|||
Db   1044 TTTTATCGAGGTGTCAAGCGAATTGGTTATGCTAGACTTCATCGCTCTACTATAACAGAC 1103

Qy    636 SerGluIleGlnMetAlaLysPheArgIleProAspAspProThrAsnTyrArgAspAsn 655
          :::|||:::|||      ||||||   :::      |||
Db   1104 ACCGAGCTTCAAACCGGAAAGTTTTTGGTTACCAAAGATTTA--------------GGT 1148

Qy    656 GlnLysValValIleAspHisArgGluValSerGluLysIleHisPheAsnProArgPhe 675
          :::   |||   ||||||         |||   |||         :::         |||
Db   1149 AGGAACGTCAGAATCGACAAGGAACACGTTCCTGAATCACACAGGTATCCATCATTATTT 1208

Qy    676 ---GlySerTyrLysGluGlyHisAsnTyrGluAsnAsnHisHisPheHisMetAsnThr 694
             |||:::         ||||||   |||:::||||||         |||
Db   1209 GAAGGAACGGGTGTTAATGGACATTACTATGACAATAATTTAGTTGTTCAC------GAG 1262

Qy    695 ProLysTyrPhe 698
          |||   |||:::
Db   1263 CCGTGGTACTAT 1274

RESULT 5
```

```
D89130
LOCUS       D89130                1234 bp    mRNA    linear   PLN 13-MAR-1998
DEFINITION  Schizosaccharomyces pombe mRNA, partial cds, clone: SY 0503.
ACCESSION   D89130
VERSION     D89130.1  GI:1749467
KEYWORDS    .
SOURCE      Schizosaccharomyces pombe (strain:PR745) cDNA to mRNA,
            clone_lib:library of H. Nojima clone:SY 0503.
  ORGANISM  Schizosaccharomyces pombe
            Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
            Schizosaccharomycetales; Schizosaccharomycetaceae;
            Schizosaccharomyces.
REFERENCE   1  (bases 1 to 1234)
  AUTHORS   Yoshioka,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-NOV-1996) Sachiyo Yoshioka, Tsukita Cell Axis Project
            ERATO JST, Kyoto Research Park; 17 Chudouji Minamimachi,
            Shimokyo-ku, Kyoto, Kyoto 600, Japan
            (E-mail:syoshi@cell.tsukita.jst.go.jp, Tel:+81-75-315-7913,
            Fax:+81-75-315-6420)
REFERENCE   2  (sites)
  AUTHORS   Yoshioka,S., Kato,K., Nakai,K., Okayama,H. and Nojima,H.
  TITLE     Identification of open reading frames in Schizosaccharomyces pombe
            cDNAs
  JOURNAL   DNA Res. 4 (6), 363-369 (1997)
  MEDLINE   98162722
FEATURES             Location/Qualifiers
     source          1. .1234
                     /organism="Schizosaccharomyces pombe"
                     /strain="PR745"
                     /db_xref="taxon:4896"
                     /clone="SY 0503"
                     /clone_lib="library of H. Nojima"
     CDS             <1. .1207
                     /note="similar to Saccharomyce cerevisiae hypothetical
                     107.9KD protein in POL4-SRD1 intergenic region, SWISS-PROT
                     Accession Number P25618"
                     /codon_start=2
                     /protein_id="BAA13792.1"
                     /db_xref="GI:1749468"
                     /translation="IYSFASIRISSFYSPAKVWGGAFLVYILYSLAHVWVVAYEFVPG
                     GPILRERTSYILIFIGWNLAALVPAYSGESKEPNKADSSVVDIKQSDSSYRRRSFKKS
                     LLTGFCLALMALKFAIQNMPPYDYTPYHPNEKLFTAGIWTIHFGLDNFMYASENRIRD
                     AVRDMELDVFGLLESDTQRLIMGFRDLTQVLAHDLGMYADYGPGPDKHTWGAALLSKF
                     PIVNSTHHLLPSPQGELAPAIHATLDVYGELIDVVVSHNGQYESQLDRRLQSTELARI
                     MRESPRPLVFLGYVVSNVGQEPQTILTRDTGMLDIEPADYDRWCQYIFYRGVKRIGYA
                     RLHRSTITDTELQTGKFLVTKDLGRNVRIDKEHVPESHRYPSLFEGTGVNGHYYDNNL
                     VVHEPWYYD"
BASE COUNT      333 a    253 c    253 g    395 t
ORIGIN

Alignment Scores:
Pred. No.:              4.16e-42        Length:         1234
Score:                  625.50          Matches:        145
Percent Similarity:     50.98%          Conservative:   63
Best Local Similarity:  35.54%          Mismatches:     169
Query Match:            16.81%          Indels:         31
DB:                     8               Gaps:           7

US-09-809-638-2 (1-699) x D89130 (1-1234)

Qy    306 IleAsnSerGlyThrAsnProGlyLysThrMetThrIleAlaMetIlePheTyrLeuLeu 325
          |||:::|||      :::|||  |||            |||   :::   |||:::|||
Db     26 ATTTCCTCCTTCTATTCTCCTGCGAAGGTTTGGGGTGGCGCTTTTTTGGTTTACATCCTT 85

Qy    326 GluIlePhePheCysAlaTrpCysThrAlaPheLysPheValProGlyGlyValTyrAla 345
                        |||      |||::::::|||||||||||||||
Db     86 TACTCATTAGCTCATGTTTGGGTCGTAGCCTACGAGTTTGTTCCAGGTGGACCCATTCTT 145

Qy    346 ArgGluArgSerAspValLeuLeuGlyThrMetMetLeuIleIleGlyLeuAsnMetLeu 365
          |||||||||:::      :::|||           :::   ||||||   |||:::
Db    146 CGAGAGAGAACTTCTTACATTCTA-----------ATTTTCATTGGTTGGAACCTTGCA 193
```

```
Qy     366 Phe----------------------------GlyProLysLysAsnLeuAspLeuLeu 375
                                          |||   |||           :::
Db     194 GCTTTAGTACCTGCGTATTCGGGTGAGTCCAAAGAACCCAACAAAGCAGACTCATCTGTA 253

Qy     376 LeuGlnThrLysAsnSerSerLysValLeuPheArgLysSerGluLysTyrMetLysLeu 395
           :::     |||   |||               |||:::|||   |||
Db     254 GTAGATATCAAACAATCAGATTCTTCTTATCGTAGAAGAAGTTTCAAAAAATCA------ 307

Qy     396 PheLeuTrpLeuLeuValGlyValGlyLeuLeuGlyLeuGlyLeuArgHisLysAlaTyr 415
                 ||||||   |||     |||      :::   |||:::
Db     308 ---------CTATTGACAGGTTTTTGTCTGGCTCTTATGGCCCTAAAATTTGCCATTCAG 358

Qy     416 GluArgLysLeuGlyLysValAlaPro------------ThrLysGluValSerAlaAla 431
                               |||                |||     :::|||
Db     359 AATATGCCTCCGTATGACTATACACCTTACCATCCAAATGAAAAGCTTTTCACTGCTGGT 418

Qy     432 IleTrpProPheArgPheGlyTyrAspAsnGluGlyTrpSerSerLeuGluArgSerAla 451
           ||||||         ||||||  ||||||      ::::::|||      |||
Db     419 ATTTGGACTATTCACTTTGGCTTAGACAATTTCATGTATGCAAGTGAGAATCGTATACGC 478

Qy     452 HisLeuLeuAsnGluThrGlyAlaAspPheIleThrIleLeuGluSerAspAlaSerLys 471
             :::  :::         |||        :::||||||||||||      :::
Db     479 GATGCTGTTCGAGATATGGAACTTGATGTCTTTGGTCTACTGGAATCTGACACACAACGT 538

Qy     472 ProTyrMetGlyAsnAsnAspLeuThrMetTrpLeuGlyGluLysLeuGlyPheTyrThr 491
             ||||||     |||||||||     |||          ||||||   |||
Db     539 TTAATTATGGGATTTCGAGATTTAACCCAGGTTTTAGCACATGATCTAGGAATGTATGCG 598

Qy     492 AspPheGlyProSerThrArgTyrHisThrTrpGlyIleMetAlaLeuSerArgTyrPro 511
           |||:::||||||             ||||||||||||         ||||||::::::|||
Db     599 GATTATGGCCCAGGCCCTGATAAACATACTTGGGGCGCTGCGTTACTCTCCAAATTTCCT 658

Qy     512 IleValLysSerGluHisHisLeuLeuProSerProGluGlyGluIleAlaProAlaIle 531
           ||||||  |||  ||||||||||||||||||:::||||||:::||||||||||||
Db     659 ATTGTCAACTCGACTCACCATTTGTTACCTTCTCCTCAAGGAGAACTTGCTCCTGCTATT 718

Qy     532 ThrLeuThrValAsnIleSerGlyLysLeuValAspPheValValThrHisPheGlyAsn 551
                |||:::::::::   |||:::|||:::|||   ||||||:::|||   |||
Db     719 CATGCAACACTAGATGTCTACGGAGAGCTAATAGATGTTGTGGTTTCTCACAATGGTCAA 778

Qy     552 HisGluAspAspLeuAspArgLysLeuGlnAlaIleAlaValSerLysLeuLeuLysSer 571
           :::|||      |||||||||:::||||||:::      :::::::::::::::::
Db     779 TATGAAAGTCAGTTGGATCGACGTCTTCAGAGCACTGAGTTGGCTCGAATTATGCGCGAG 838

Qy     572 SerSerAsnGlnValIlePheLeuGlyTyrIleThrSerAlaProGlySerArgAspTyr 591
           |||        ::::::||||||||||||:::   |||      |||
Db     839 AGTCCAAGGCCTCTAGTGTTTCTCGGATATGTTGTTTCAAACGTTGGCCAAGAACCCCAA 898

Qy     592 LeuGlnLeuThrGluHisGlyAsnValLysAspIleAspSerThrAspHisAspArgTrp 611
                |||||          :::   ||||||:::      |||:::||||||||
Db     899 ACTATTCTAACGCGAGACACGGGAATGTTAGACATTGAACCTGCTGACTATGATCGTTGG 958

Qy     612 CysGluTyrIleMetTyrArgGlyLeuIleArgLeuGlyTyrAlaArgIleSerHisAla 631
           |||:::||||||   |||||||||:::   |||:::|||||||||||||:::      :::
Db     959 TGTCAGTATATATTTTATCGAGGTGTCAAGCGAATTGGTTATGCTAGACTTCATCGCTCT 1018

Qy     632 GluLeuSerAspSerGluIleGlnMetAlaLysPheArgIleProAspAspProThrAsn 651
              ::::::|||:::|||:::|||      ||||||   :::      |||
Db    1019 ACTATAACAGACACCGAGCTTCAAACCGGAAAGTTTTTGGTTACCAAAGATTTA------ 1072

Qy     652 TyrArgAspAsnGlnLysValValIleAspHisArgGluValSerGluLysIleHisPhe 671
                    :::   |||   ||||||         |||   |||      :::
Db    1073 ---------GGTAGGAACGTCAGAATCGACAAGGAACACGTTCCTGAATCACACAGGTAT 1123

Qy     672 AsnProArgPhe---GlySerTyrLysGluGlyHisAsnTyrGluAsnAsnHisHisPhe 690
                  |||   |||:::         ||||||   |||:::||||||
Db    1124 CCATCATTATTTGAAGGAACGGGTGTTAATGGACATTACTATGACAATAATTTAGTTGTT 1183

Qy     691 HisMetAsnThrProLysTyrPhe 698
           |||         |||   |||:::
Db    1184 CAC------GAGCCGTGGTACTAT 1201

RESULT 6
```

```
CNS06JWV
LOCUS       CNS06JWV                995 bp    DNA     linear   STS 10-JAN-2001
DEFINITION  T3 end of clone XAS0AA002B08 of library XAS0AA from strain CLIB 533
            of Saccharomyces bayanus, sequence tagged site.
ACCESSION   AL402149
VERSION     AL402149.1  GI:12160780
KEYWORDS    STS.
SOURCE      Saccharomyces bayanus.
  ORGANISM  Saccharomyces bayanus
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE   1  (bases 1 to 995)
  AUTHORS   Bon,E., Neuveglise,C., Casaregola,S., Artiguenave,F., Wincker,P.,
            Aigle,M. and Durrens,P.
  TITLE     Genomic Exploration of the Hemiascomycetous Yeasts: 5.
            Saccharomyces bayanus var. uvarum
  JOURNAL   FEBS Lett. 487 (1), 37-41 (2000)
   PUBMED   11152880
REFERENCE   2  (bases 1 to 995)
  AUTHORS   Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
            Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,
            de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B.,
            Malpertuy,A., Neuveglise,C., Ozier-Kalogeropoulos,O., Potier,S.,
            Saurin,W., Tekaia,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,
            Wincker,P. and Weissenbach,J.
  TITLE     Genomic Exploration of the Hemiascomycetous Yeasts: 1. A set of
            yeast species for molecular evolution studies(1)
  JOURNAL   FEBS Lett. 487 (1), 3-12 (2000)
   PUBMED   11152876
REFERENCE   3  (bases 1 to 995)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-SEP-2000) Genoscope - Centre National de Sequencage,
            2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
            seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT     This STS is part of a random genomic sequencing program of thirteen
            yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
            exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
            Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
            lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
            angusta, Debaryomyces hansenii var.  hansenii, Pichia sorbitophila,
            Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
            5 kb were prepared and both extremities were sequenced. See
            keywords for description of this sequence and for the sequence of
            the other extremity of this insert.
FEATURES             Location/Qualifiers
     source          1. .995
                     /organism="Saccharomyces bayanus"
                     /strain="CLIB 533"
                     /variety="uvarum"
                     /db_xref="taxon:4931"
                     /clone="XAS0AA002B08"
                     /clone_lib="XAS0AA"
                     /note="end : T3"
     misc_feature    <3. .>958
                     /note="similar to Saccharomyces cerevisiae ORF YGR017c [
                     similarity to hypothetical S.pombe protein ]
                     1 putative frameshift(s)"
                     /evidence=not_experimental
BASE COUNT      276 a    190 c    210 g    318 t      1 others
ORIGIN

Alignment Scores:
Pred. No.:                1.53e-31        Length:          995
Score:                    495.00          Matches:         136
Percent Similarity:       50.55%          Conservative:    48
Best Local Similarity:    37.36%          Mismatches:      136
Query Match:              13.31%          Indels:          47
DB:                       11              Gaps:            10

US-09-809-638-2 (1-699) x CNS06JWV (1-995)

Qy     274 GlyLeuLeuTyrLeuHisThrTrpAlaAlaAlaValSerGlyCysValPheAlaIlePhe 293
```

```
Db      6 GGTTTATTGTAC-----------GCCATTGCTATGCTTTGGTTAGTTCCTTCATACTTT 53

Qy    294 ThrAla--------------SerMetTrpProGlnThrLeuGlyHisLeuIleAsnSer 308
Db     54 TCCGTACTAGGTCAAGTACAAAGTATATGGGTTTACGTTTTC----------------- 95

Qy    309 GlyThrAsnProGlyLysThrMetThrIleAlaMetIlePheTyrLeuLeuGluIlePhe 328
Db     96 -----------------TCATTTGCTGTCTATGTTATTTTTGTCTTGGCTCATGTT--- 134

Qy    329 PheCysAlaTrpCysThrAlaPheLysPheValProGlyGlyValTyrAlaArgGluArg 348
Db    135 ---------TGGGTCGTCGCATATGCATTCGTTCCATTAGGTTGGTTACTGAGGGAGAGA 185

Qy    349 SerAspValLeuLeuGlyThrMetMetLeuIleIleGlyLeuAsnMetLeuPheGlyPro 368
Db    186 ATCGAGGCTATCCTTGCAGTTTCCTCGACGTTTATTATCCTTGGTATGTTGGTCAACAAT 245

Qy    369 LysLysAsnLeuAspLeuLeuLeuGlnThrLysAsnSerSerLysValLeuPheArgLys 388
Db    246 AACTTCAACCTTCGACTAATA------------AACGTAAGCAAGAAATTATGC------ 287

Qy    389 SerGluLysTyrMetLysLeuPheLeuTrpLeuLeuValGlyValGlyLeuLeuGlyLeu 408
Db    288 ------ACTTATATTTCTTTGTTT---------------GCAGTAGCTCTACTAGCATTG 326

Qy    409 GlyLeuArgHisLysAlaTyrGluArgLysLeuGlyLysValAlaPro------------ 424
Db    327 ACCGCTAGATTTACATACGAAATTAGACCAACCGGAATTCCCCAACCTTATCATCCTGAT 386

Qy    425 ThrLysGluValSerAlaAlaIleTrpProPheArgPheGlyTyrAspAsnGluGlyTrp 444
Db    387 TCTCAATTGATCTCTGCTGGTATTTGGACCATCCATTTTGGTCTTGATAATGATATGTGG 446

Qy    445 SerSerLeuGluArgSerAlaHisLeuLeuAsnGluThrGlyAlaAspPheIleThrIle 464
Db    447 GCATCTGAAGACAGAATGATCAACTTAATAAAAGAAATGGAGTTGGATGTTGTAGGATTA 506

Qy    465 LeuGluSerAspAlaSerLysProTyrMetGlyAsnAsnAspLeuThrMetTrpLeuGly 484
Db    507 CTAGAAACAGATACGCAAAGAATTATTATGGGTAACAGGGATCTAACCAGTAAACTAGCC 566

Qy    485 GluLysLeuGlyPheTyrThrAspPheGlyProSerThrArgTyrHisThrTrpGlyIle 504
Db    567 CATGATCTGAATATGTATGCTGATTTTGGCCCGGGTCCAAACAAACACACATGGGGCTGT 626

Qy    505 MetAlaLeuSerArgTyrProIleValLysSerGluHisHisLeuLeuProSerProGlu 524
Db    627 ATTCTTCTTTCCAAATTCCCTATCATTAATGCAACGCACCATCTATTGCCATCCCCTGTG 686

Qy    525 GlyGluIleAlaProAlaIleThrLeuThrVal---AsnIleSerGlyLysLeuValAsp 543
Db    687 GGAGAGCTTGCGCCGGCCATTCACGCCACTCTTCAAACGTATAATGACACGTTAGTTGAT 746

Qy    544 PheValValThrHisPheGlyAsnHisGluAspAspLeuAspArgLysLeuGlnAlaIle 563
Db    747 GTCTTTGTCTTCCATAGTGGACAAGAGGAAGATGAAGAAGATAGAAGATTACAAAGTAAT 806

Qy    564 AlaValSerLysLeuLeuLysSerSerSerAsnGlnValIlePheLeuGlyTyrIleThr 583
Db    807 TACATGGCTAAATTGATGGGAAATTCGACCCGACCAGCTATTTTATTAAGTTACTTGGTG 866

Qy    584 SerAlaProGlySerArgAspTyr---LeuGlnLeuThrGluHisGlyAsnValLysAsp 602
Db    867 GTTGACCCGGGCGAAGGTAACTACAATACGCATGTTAGTGAACT-TCTGGCATGCATGAC 925

Qy    603 IleAspSerThrAspHisAspArgTrpCysGluTyrIleMetTyrArgGlyLeuIleArg 622
Db    926 ATTGACCCTACTGACGATGATAGGTGGTGTGA-TATATCTTGTATAAAGATTTGAGAAGA 984

Qy    623 LeuGlyTyrAla 626
```

```
Db    985 CA-GGKTATGCA 995

RESULT 7
AC020593
LOCUS       AC020593              184206 bp    DNA     linear   PRI 07-NOV-2001
DEFINITION  Homo sapiens BAC clone RP11-317G22 from 4, complete sequence.
ACCESSION   AC020593
VERSION     AC020593.6  GI:14589709
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 184206)
  AUTHORS   Sulston,J.E. and Waterston,R.
  TITLE     Toward a complete human genome sequence
  JOURNAL   Genome Res. 8 (11), 1097-1108 (1998)
  MEDLINE   99063792
REFERENCE   2  (bases 1 to 184206)
  AUTHORS   Swearengen,S. and Cotton,M.
  TITLE     The sequence of Homo sapiens BAC clone RP11-317G22
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 184206)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-JAN-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   4  (bases 1 to 184206)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-JUL-2001) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   5  (bases 1 to 184206)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-NOV-2001) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     On Jul 3, 2001 this sequence version replaced gi:12057004.
            -------------- Genome Center
                Center: Washington University Genome Sequencing Center
                Center code: WUGSC
                Web site: http://genome.wustl.edu/gsc
                Contact: sapiens@watson.wustl.edu
            -------------- Summary Statistics
                Center project name: H_NH0317G22
            --------------.

            NOTICE:  This sequence may not represent the entire insert of this
            clone.  It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one subclone; and the assembly was confirmed by
            restriction digest.

            MAPPING INFORMATION:
            Mapping information for this clone was provided by Dr. John D.
            McPherson, Department of Genetics, Washington University, St. Louis
            MO.  For additional information about the map position of this
            sequence, see http://genome.wustl.edu/gsc

            SOURCE INFORMATION:
            The RPCI-11 human BAC library was made from the blood of one male
            donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E.,
            Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved
            approach for construction of bacterial artificial chromosome
```

libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (http://bacpac.med.buffalo.edu)
VECTOR: pBACe3.6
NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the left is AC011956. Actual start of this clone is at base position 1 of RP11-317G22; actual end is at base position 184206 of RP11-317G22.

The sequence fidelity of RP11-317G22 between bases 17725 to 17956 can not be guaranteed due to an unresolved dinucleotide repeat. The sequence fidelity of RP11-317G22 between bases 86965 to 87015 can not be guaranteed due to an unresolved homopolymeric run.

FEATURES Location/Qualifiers
source 1. .184206
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="4"
/map="4"
/clone="RP11-317G22"
/clone_lib="RPCI-11"
repeat_region 5. .70
/rpt_family="L2"
misc_feature 252. .374
/note="match to EST BG169511 (NID:g12676214)"
misc_feature 252. .374
/note="match to EST AI088656 (NID:g3427715) qb14a11.x1"
misc_feature 252. .374
/note="match to EST BG214746 (NID:g13740767)"
misc_feature 252. .374
/note="match to EST BF055447 (NID:g10809343)"
misc_feature 252. .374
/note="match to EST AI077951 (NID:g3412359) oy39a10.s1"
misc_feature 252. .374
/note="match to EST AW006753 (NID:g5855531) wt07a02.x1"
misc_feature 252. .374
/note="match to EST BG291592 (NID:g13049843)"
misc_feature 252. .374
/note="match to EST W80356 (NID:g1391393) zh49a05.s1"
misc_feature 252. .494
/note="match to EST AA464688 (NID:g2189572) zx82a02.s1"
misc_feature 252. .374
/note="match to EST AW005009 (NID:g5853713) wv02e08.x1"
misc_feature 252. .374
/note="match to EST BG165990 (NID:g12672693)"
misc_feature 256. .374
/note="similar to Bos taurus EST AV609860 (NID:g9745530)"
misc_feature 280. .374
/note="match to EST AW371794 (NID:g6876448)"
misc_feature 347. .374
/note="match to EST AA303948 (NID:g1956281)"
misc_feature 708. .1203
/note="match to EST BI063872 (NID:g14471399)"
misc_feature 1021. .1308
/note="match to EST T06930 (NID:g318079)"
misc_feature 1119. .1143
/note="match to EST W68728 (NID:g1377608) zd37c01.s1"
misc_feature 1119. .1143
/note="match to EST AI220135 (NID:g3802338) qi44g05.x1"
misc_feature 1119. .1143
/note="match to EST AI271747 (NID:g3890914) qj77b10.x1"
misc_feature 1119. .1143
/note="match to EST BG167015 (NID:g12673728)"
misc_feature 1183. .1538
/note="match to EST BG546183 (NID:g13544848)"
repeat_region 1319. .1350
/rpt_family="AT_rich"
repeat_region 1349. .1466
/rpt_family="Alu"
misc_feature 1490. .1538
/note="match to EST AA847380 (NID:g2933898) of07d05.s1"
misc_feature 1490. .1538
/note="match to EST AI220135 (NID:g3802338) qi44g05.x1"
misc_feature 1490. .1538
/note="match to EST W68728 (NID:g1377608) zd37c01.s1"
misc_feature 1490. .1538
/note="match to EST BG167015 (NID:g12673728)"
misc_feature 1490. .1538
/note="match to EST AI271747 (NID:g3890914) qj77b10.x1"
misc_feature 1491. .1538
/note="match to EST W80356 (NID:g1391393) zh49a05.s1"
misc_feature 1491. .1538
/note="match to EST AI088656 (NID:g3427715) qb14a11.x1"
misc_feature 1491. .1508
/note="match to EST BG291592 (NID:g13049843)"
misc_feature 1491. .1538
/note="similar to Homo sapiens EST BG207366 (NID:g13729053)"
misc_feature 1491. .1538
/note="match to EST BG165990 (NID:g12672693)"
misc_feature 1491. .1538
/note="match to EST AA303948 (NID:g1956281)"
misc_feature 1491. .1538
/note="match to EST AW371794 (NID:g6876448)"
misc_feature 1491. .1538
/note="match to EST AW005009 (NID:g5853713) wv02e08.x1"
misc_feature 1491. .1538
/note="similar to Bos taurus EST AV609860 (NID:g9745530)"
misc_feature 1491. .1538
/note="match to EST AI077951 (NID:g3412359) oy39a10.s1"
misc_feature 1491. .1538
/note="match to EST BG214746 (NID:g13740767)"
misc_feature 1491. .1538
/note="match to EST AW006753 (NID:g5855531) wt07a02.x1"
misc_feature 1491. .1538
/note="match to EST BG169511 (NID:g12676214)"
misc_feature 1491. .1546
/note="match to EST BF055447 (NID:g10809343)"
repeat_region 1561. .1606
/rpt_family="AT_rich"
repeat_region 1696. .1986
/rpt_family="Alu"
repeat_region 2175. .2217
/rpt_family="L1"
repeat_region 2311. .2780
/rpt_family="MaLR"
repeat_region 2828. .3176
/rpt_family="MaLR"
repeat_region 3217. .3544
/rpt_family="Alu"
repeat_region 3617. .3648
/rpt_family="AT_rich"
repeat_region 3793. .4078
/rpt_family="Alu"
repeat_region 4138. .4267
/rpt_family="Alu"
repeat_region 4462. .4655
/rpt_family="MER1_type"
misc_feature 4639. .4688
/note="match to EST AW005009 (NID:g5853713) wv02e08.x1"

Alignment Scores:

Pred. No.:	1.43e-26	Length:	184206
Score:	471.00	Matches:	88
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	12.66%	Indels:	0
DB:	9	Gaps:	0

US-09-809-638-2 (1-699) x AC020593 (1-184206)

Qy 267 ThrGlyThrAlaSerAlaAlaGlyLeuLeuTyrLeuHisThrTrpAlaAlaAlaValSer 286
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 111216 ACAGGAACAGCTTCAGCTGCGGGGCTCCTTTACCTGCACACATGGGCAGCTGCTGTGTCT 111275

```
Qy     287 GlyCysValPheAlaIlePheThrAlaSerMetTrpProGlnThrLeuGlyHisLeuIle 306
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 111276 GGCTGTGTCTTCGCCATCTTTACTGCATCCATGTGGCCCCAAACACTTGGACACCTTATT 111335

Qy     307 AsnSerGlyThrAsnProGlyLysThrMetThrIleAlaMetIlePheTyrLeuLeuGlu 326
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 111336 AACTCAGGGACAAACCCTGGGAAAACCATGACCATTGCCATGATATTTTATCTTCTAGAA 111395

Qy     327 IlePhePheCysAlaTrpCysThrAlaPheLysPheValProGlyGlyValTyrAlaArg 346
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 111396 ATATTTTTCTGTGCCTGGTGCACAGCTTTTAAGTTTGTCCCAGGAGGTGTCTACGCTAGA 111455

Qy     347 GluArgSerAspValLeuLeuGly 354
           |||||||||||||||||||||||||
Db 111456 GAAAGATCAGATGTGCTTTTGGGT 111479

RESULT 8
AL158816
ID   AL158816             222885 bp    DNA     linear   HTG 10-JUL-2001
DEFINITION  Homo sapiens chromosome 6 clone RP3-405B4, *** SEQUENCING IN
            PROGRESS ***, 33 unordered pieces.
ACCESSION   AL158816
VERSION     AL158816.11  GI:9943990
KEYWORDS    HTG; HTGS_PHASE1; HTGS_CANCELLED.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 222885)
  AUTHORS   Sims,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Aug 29, 2000 this sequence version replaced gi:9926474.
            -------------- Genome Center
            Center: Sanger Centre
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            -------------- Project Information
            Center project name: dJ405B4
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Sequencing vector: plasmid; L08752; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Consensus quality: 204390 bases at least Q40
            Consensus quality: 211527 bases at least Q30
            Consensus quality: 215491 bases at least Q20
            Insert size: 219685; sum-of-contigs
            Insert size: 126549; 14.0% error; agarose-fp
            Quality coverage: 3.63x in Q20 bases; sum-of-contigs Quality
            coverage: 6.72x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 33 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     3599: contig of 3599 bp in length
            *     3600 3699: gap of      100 bp
            *     3700    12818: contig of 9119 bp in length
            *    12819 12918: gap of      100 bp
            *    12919    16398: contig of 3480 bp in length
            *    16399 16498: gap of      100 bp
            *    16499    18590: contig of 2092 bp in length
            *    18591 18690: gap of      100 bp
            *    18691    22372: contig of 3682 bp in length
            *    22373 22472: gap of      100 bp
            *    22473    26905: contig of 4433 bp in length
            *    26906 27005: gap of      100 bp
            *    27006    29581: contig of 2576 bp in length
            *    29582 29681: gap of      100 bp
            *    29682    33446: contig of 3765 bp in length
            *    33447 33546: gap of      100 bp
            *    33547    48536: contig of 14990 bp in length
            *    48537 48636: gap of      100 bp
            *    48637    55599: contig of 6963 bp in length
            *    55600 55699: gap of      100 bp
            *    55700    65124: contig of 9425 bp in length
            *    65125 65224: gap of      100 bp
            *    65225    69090: contig of 3866 bp in length
            *    69091 69190: gap of      100 bp
            *    69191    73894: contig of 4704 bp in length
            *    73895 73994: gap of      100 bp
            *    73995    82665: contig of 8671 bp in length
            *    82666 82765: gap of      100 bp
            *    82766    87239: contig of 4474 bp in length
            *    87240 87339: gap of      100 bp
            *    87340    93671: contig of 6332 bp in length
            *    93672 93771: gap of      100 bp
            *    93772    97550: contig of 3779 bp in length
            *    97551 97650: gap of      100 bp
            *    97651   100101: contig of 2451 bp in length
            *   100102 100201: gap of      100 bp
            *   100202   103416: contig of 3215 bp in length
            *   103417 103516: gap of      100 bp
            *   103517   107564: contig of 4048 bp in length
            *   107565 107664: gap of      100 bp
            *   107665   131778: contig of 24114 bp in length
            *   131779 131878: gap of      100 bp
            *   131879   135261: contig of 3383 bp in length
            *   135262 135361: gap of      100 bp
            *   135362   137913: contig of 2552 bp in length
            *   137914 138013: gap of      100 bp
            *   138014   140949: contig of 2936 bp in length
            *   140950 141049: gap of      100 bp
            *   141050   156466: contig of 15417 bp in length
            *   156467 156566: gap of      100 bp
            *   156567   164836: contig of 8270 bp in length
            *   164837 164936: gap of      100 bp
            *   164937   168950: contig of 4014 bp in length
            *   168951 169050: gap of      100 bp
            *   169051   173724: contig of 4674 bp in length
            *   173725 173824: gap of      100 bp
            *   173825   179991: contig of 6167 bp in length
            *   179992 180091: gap of      100 bp
            *   180092   190208: contig of 10117 bp in length
            *   190209 190308: gap of      100 bp
            *   190309   193471: contig of 3163 bp in length
            *   193472 193571: gap of      100 bp
            *   193572   197829: contig of 4258 bp in length
            *   197830 197929: gap of      100 bp
            *   197930   222885: contig of 24956 bp in length.
FEATURES             Location/Qualifiers
     source          1. .222885
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /clone="RP3-405B4"
                     /clone_lib="RPCI-3"
     misc_feature    1. .3599
                     /note="assembly_fragment:00027"
     misc_feature    3700. .12818
                     /note="assembly_fragment:00056"
     misc_feature    12919. .16398
                     /note="assembly_fragment:00093"
     misc_feature    16499. .18590
                     /note="assembly_fragment:00125"
     misc_feature    18691. .22372
                     /note="assembly_fragment:00140"
     misc_feature    22473. .26905
                     /note="assembly_fragment:00167"
```

```
     misc_feature    27006. .29581
                     /note="assembly_fragment:00175"
     misc_feature    29682. .33446
                     /note="assembly_fragment:00178"
     misc_feature    33547. .48536
                     /note="assembly_fragment:00416"
     misc_feature    48637. .55599
                     /note="assembly_fragment:00542"
     misc_feature    55700. .65124
                     /note="assembly_fragment:00565"
     misc_feature    65225. .69090
                     /note="assembly_fragment:00583"
     misc_feature    69191. .73894
                     /note="assembly_fragment:00649"
     misc_feature    73995. .82665
                     /note="assembly_fragment:00709"
     misc_feature    82766. .87239
                     /note="assembly_fragment:00789"
     misc_feature    87340. .93671
                     /note="assembly_fragment:00874"
     misc_feature    93772. .97550
                     /note="assembly_fragment:00924"
     misc_feature    97651. .100101
                     /note="assembly_fragment:01052"
     misc_feature    100202. .103416
                     /note="assembly_fragment:01082"
     misc_feature    103517. .107564
                     /note="assembly_fragment:01216"
     misc_feature    107665. .131778
                     /note="assembly_fragment:01234"
     misc_feature    131879. .135261
                     /note="assembly_fragment:01280"
     misc_feature    135362. .137913
                     /note="assembly_fragment:01337"
     misc_feature    138014. .140949
                     /note="assembly_fragment:01643"
     misc_feature    141050. .156466
                     /note="assembly_fragment:01660"
     misc_feature    156567. .164836
                     /note="assembly_fragment:01690"
     misc_feature    164937. .168950
                     /note="assembly_fragment:01692.0"
     misc_feature    169051. .173724
                     /note="assembly_fragment:01748"
     misc_feature    173825. .179991
                     /note="assembly_fragment:02061"
     misc_feature    180092. .190208
                     /note="assembly_fragment:02140"
     misc_feature    190309. .193471
                     /note="assembly_fragment:02150"
     misc_feature    193572. .197829
                     /note="assembly_fragment:02181"
     misc_feature    197930. .222885
                     /note="assembly_fragment:00834
                     clone_end:T7
                     vector_side:right"
BASE COUNT    69666 a  41364 c  44947 g  63682 t   3226 others
ORIGIN

Alignment Scores:
Pred. No.:              1.97e-11        Length:         222885
Score:                  287.50          Matches:        63
Percent Similarity:     70.30%          Conservative:   8
Best Local Similarity:  62.38%          Mismatches:     15
Query Match:            7.73%           Indels:         15
DB:                     2               Gaps:           3

US-09-809-638-2 (1-699) x AL158816 (1-222885)

Qy     606 ThrAspHisAspArgTrpCysGluTyrIleMetTyrArgGlyLeuIleArgLeuGlyTyr 625
           |||:::|||     :::|||                          ||||||||||||
Db 157651 ACAAACCATTTCTGTTTCTGC----------------------TACAGGTTGGGTTAT 157686

Qy     626 AlaArgIleSerHisAlaGluLeuSerAspSerGluIleGlnMetAlaLysPheArgIle 645
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 157687 GCAAGAATCTCCCATGCTGAACTGAGTGATTCAGAAATTCAGATGGCAAAATTTAGGATC 157746

Qy     646 ProAspAspProThrAsnTyrArgAspAsnGlnLysValValIleAspHisArgGluVal 665
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 157747 CCTGATGACCCCACTAATTATAGAGACAACCAGAAAGTGGTCATAGACCACAGAGAAGTT 157806

Qy     666 SerGluLysIleHisPheAsnProArg----------------PheGlySerTyrLys 679
           |||||||||||||||||||||||||                ||||||:::
Db 157807 TCTGAGAAAATTCATTTTAATCCCAGGTGAGTTCCTTTATGCTGTAGTTCCAGTTATGAG 157866

Qy     680 GluGlyHis---AsnTyrGluAsnAsnHisHisPheHisMetAsnThrProLysTyrPhe 698
           |||   :::   |||      ::::::      |||:::         ||||||
Db 157867 CTACTGCATCCCTCTATGGAACCTACATATTACCTCAAAATGAGCTAGAAATTATATTTT 157926

Qy     699 Leu 699
           :::
Db 157927 GTC 157929

RESULT 9
AL158816/c
LOCUS       AL158816              222885 bp    DNA     linear   HTG 10-JUL-2001
DEFINITION  Homo sapiens chromosome 6 clone RP3-405B4, *** SEQUENCING IN
            PROGRESS ***, 33 unordered pieces.
ACCESSION   AL158816
VERSION     AL158816.11  GI:9943990
KEYWORDS    HTG; HTGS_PHASE1; HTGS_CANCELLED.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 222885)
  AUTHORS   Sims,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Aug 29, 2000 this sequence version replaced gi:9926474.
            -------------- Genome Center
            Center: Sanger Centre
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            -------------- Project Information
            Center project name: dJ405B4
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Sequencing vector: plasmid; L08752; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Consensus quality: 204390 bases at least Q40
            Consensus quality: 211527 bases at least Q30
            Consensus quality: 215491 bases at least Q20
            Insert size: 219685; sum-of-contigs
            Insert size: 126549; 14.0% error; agarose-fp
            Quality coverage: 3.63x in Q20 bases; sum-of-contigs Quality
            coverage: 6.72x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 33 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1     3599: contig of 3599 bp in length
            *    3600 3699: gap of      100 bp
            *    3700    12818: contig of 9119 bp in length
            *   12819 12918: gap of      100 bp
            *   12919    16398: contig of 3480 bp in length
            *   16399 16498: gap of      100 bp
            *   16499    18590: contig of 2092 bp in length
```

```
            *    18591 18690: gap of      100 bp
            *    18691    22372: contig of 3682 bp in length
            *    22373 22472: gap of      100 bp
            *    22473    26905: contig of 4433 bp in length
            *    26906 27005: gap of      100 bp
            *    27006    29581: contig of 2576 bp in length
            *    29582 29681: gap of      100 bp
            *    29682    33446: contig of 3765 bp in length
            *    33447 33546: gap of      100 bp
            *    33547    48536: contig of 14990 bp in length
            *    48537 48636: gap of      100 bp
            *    48637    55599: contig of 6963 bp in length
            *    55600 55699: gap of      100 bp
            *    55700    65124: contig of 9425 bp in length
            *    65125 65224: gap of      100 bp
            *    65225    69090: contig of 3866 bp in length
            *    69091 69190: gap of      100 bp
            *    69191    73894: contig of 4704 bp in length
            *    73895 73994: gap of      100 bp
            *    73995    82665: contig of 8671 bp in length
            *    82666 82765: gap of      100 bp
            *    82766    87239: contig of 4474 bp in length
            *    87240 87339: gap of      100 bp
            *    87340    93671: contig of 6332 bp in length
            *    93672 93771: gap of      100 bp
            *    93772    97550: contig of 3779 bp in length
            *    97551 97650: gap of      100 bp
            *    97651   100101: contig of 2451 bp in length
            *   100102 100201: gap of      100 bp
            *   100202   103416: contig of 3215 bp in length
            *   103417 103516: gap of      100 bp
            *   103517   107564: contig of 4048 bp in length
            *   107565 107664: gap of      100 bp
            *   107665   131778: contig of 24114 bp in length
            *   131779 131878: gap of      100 bp
            *   131879   135261: contig of 3383 bp in length
            *   135262 135361: gap of      100 bp
            *   135362   137913: contig of 2552 bp in length
            *   137914 138013: gap of      100 bp
            *   138014   140949: contig of 2936 bp in length
            *   140950 141049: gap of      100 bp
            *   141050   156466: contig of 15417 bp in length
            *   156467 156566: gap of      100 bp
            *   156567   164836: contig of 8270 bp in length
            *   164837 164936: gap of      100 bp
            *   164937   168950: contig of 4014 bp in length
            *   168951 169050: gap of      100 bp
            *   169051   173724: contig of 4674 bp in length
            *   173725 173824: gap of      100 bp
            *   173825   179991: contig of 6167 bp in length
            *   179992 180091: gap of      100 bp
            *   180092   190208: contig of 10117 bp in length
            *   190209 190308: gap of      100 bp
            *   190309   193471: contig of 3163 bp in length
            *   193472 193571: gap of      100 bp
            *   193572   197829: contig of 4258 bp in length
            *   197830 197929: gap of      100 bp
            *   197930   222885: contig of 24956 bp in length.
FEATURES             Location/Qualifiers
     source          1. .222885
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /clone="RP3-405B4"
                     /clone_lib="RPCI-3"
     misc_feature    1. .3599
                     /note="assembly_fragment:00027"
     misc_feature    3700. .12818
                     /note="assembly_fragment:00056"
     misc_feature    12919. .16398
                     /note="assembly_fragment:00093"
     misc_feature    16499. .18590
                     /note="assembly_fragment:00125"
     misc_feature    18691. .22372
                     /note="assembly_fragment:00140"
     misc_feature    22473. .26905
                     /note="assembly_fragment:00167"
     misc_feature    27006. .29581
                     /note="assembly_fragment:00175"
     misc_feature    29682. .33446
                     /note="assembly_fragment:00178"
     misc_feature    33547. .48536
                     /note="assembly_fragment:00416"
     misc_feature    48637. .55599
                     /note="assembly_fragment:00542"
     misc_feature    55700. .65124
                     /note="assembly_fragment:00565"
     misc_feature    65225. .69090
                     /note="assembly_fragment:00583"
     misc_feature    69191. .73894
                     /note="assembly_fragment:00649"
     misc_feature    73995. .82665
                     /note="assembly_fragment:00709"
     misc_feature    82766. .87239
                     /note="assembly_fragment:00789"
     misc_feature    87340. .93671
                     /note="assembly_fragment:00874"
     misc_feature    93772. .97550
                     /note="assembly_fragment:00924"
     misc_feature    97651. .100101
                     /note="assembly_fragment:01052"
     misc_feature    100202. .103416
                     /note="assembly_fragment:01082"
     misc_feature    103517. .107564
                     /note="assembly_fragment:01216"
     misc_feature    107665. .131778
                     /note="assembly_fragment:01234"
     misc_feature    131879. .135261
                     /note="assembly_fragment:01280"
     misc_feature    135362. .137913
                     /note="assembly_fragment:01337"
     misc_feature    138014. .140949
                     /note="assembly_fragment:01643"
     misc_feature    141050. .156466
                     /note="assembly_fragment:01660"
     misc_feature    156567. .164836
                     /note="assembly_fragment:01690"
     misc_feature    164937. .168950
                     /note="assembly_fragment:01692.0"
     misc_feature    169051. .173724
                     /note="assembly_fragment:01748"
     misc_feature    173825. .179991
                     /note="assembly_fragment:02061"
     misc_feature    180092. .190208
                     /note="assembly_fragment:02140"
     misc_feature    190309. .193471
                     /note="assembly_fragment:02150"
     misc_feature    193572. .197829
                     /note="assembly_fragment:02181"
     misc_feature    197930. .222885
                     /note="assembly_fragment:00834
                     clone_end:T7
                     vector_side:right"
BASE COUNT    69666 a  41364 c  44947 g  63682 t   3226 others
ORIGIN

Alignment Scores:
Pred. No.:              7.09e-10        Length:         222885
Score:                  268.50          Matches:        58
Percent Similarity:     70.93%          Conservative:   3
Best Local Similarity:  67.44%          Mismatches:     16
Query Match:            7.22%           Indels:         9
DB:                     2               Gaps:           2

US-09-809-638-2 (1-699) x AL158816 (1-222885)
```

```
Qy     457 ThrGlyAlaAspPheIleThrIleLeuGluSerAspAlaSerLysProTyrMetGlyAsn 476
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  217472 ACAGGTGCAGATTTCATAACAATTTTGGAGAGTGATGCTTCTAAGCCCTATATGGGGAAC 217413

Qy     477 AsnAspLeuThrMetTrpLeuGlyGluLysLeuGlyPheTyrThrAspPheGlyProSer 496
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  217412 AATGACTTAACCATGTGGCTAGGGGAAAAGTTGGGTTTCTATACAGACTTTGGTCCAAGC 217353

Qy     497 ThrArgTyrHisThrTrpGlyIleMetAlaLeu---------SerArgTyrProIleVal 513
           ||||||||||||||||||||   :::   |||         ||||||
Db  217352 ACAAGGTATCACACTTGGGGGTGAGTATACCTTGGGAGTTAATCCCGAAGATAGTCATGA 217293

Qy     514 LysSerGluHisHisLeuLeuProSerProGlu-----------------GlyGluIle 527
           |||:::      |||   |||                            ||||||
Db  217292 AAGGCTAGGCTGCATAGGCTTGGAGTGGCAAGTTGATATTGTAATATCTTTGGTGAATGT 217233

Qy     528 AlaProAlaIleThrLeu 533
           ||||||   :::   |||
Db  217232 GCCCCAGTGATGCTTCTT 217215
```

RESULT 10
SPAC589
LOCUS SPAC589 18342 bp DNA linear PLN 04-JAN-2001
DEFINITION S.pombe chromosome I cosmid c589.
ACCESSION AL512496
VERSION AL512496.1 GI:12043544
KEYWORDS i norganic phosphate transporter; metaxin 2; mitochondrial outer membrane protein; phosphatidylinositol metablolism; spindle integrity WD domain; ubiquitin-like protein.
SOURCE fission yeast.
ORGANISM Schizosaccharomyces pombe
Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes; Schizosaccharomycetales; Schizosaccharomycetaceae; Schizosaccharomyces.
REFERENCE 1 (bases 1 to 18342)
AUTHORS Wood,V., Rajandream,M.A., Barrell,B.G., Aert,R., Robben,J., Weltjens,I., Grymonprez,B. and Volckaert,G.
TITLE Direct Submission
JOURNAL Submitted (01-JAN-2001) European Schizosaccharomyces genome sequencing project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk and Katholieke Universiteit Leuven, Laboratory of Gene Technology, Kardinaal Mercierlaan, 9
COMMENT Notes:
Details of yeast sequencing at the Sanger Centre are available on the World Wide Web.
(URL, http://www.sanger.ac.uk/Projects/S_pombe/)
During 1995 to 1996 about 66% of S. pombe chromosome 1 was sequenced by the Sanger Centre. The sequencing of the S. pombe genome is now being continued with funding from The European Commission. Fourteen European sequencing laboratories, including the Sanger Centre, are participating in the project.
Protein coding regions (CDS) have been predicted with the help of computer analysis using the Genefinder program in PomBase (an ACEDB database) with additional predictions for the branch-acceptor sites supplied by the program Sp3splice. CAUTION: It is possible that for any individual CDS we may have underestimated or overestimated the number of introns/exons or we may not have chosen the correct splice donor/acceptor sites. CDS are numbered using the following system eg SPBC25H2.01c. SP (S. pombe), B (chromosome 2), c25H2 (cosmid name), .01 (first CDS), c (complementary strand).
The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous. The length in codons is given for each CDS.
IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions. Cosmid c589 is overlapped at the 5' end by cosmid c16A10, EMBL entry SPAC16A10, accession number Z97185, and at the 3' end by cosmid c688, EMBL entry SPAC688, accession number AL355632.

```
FEATURES             Location/Qualifiers
     source          1. .18342
                     /organism="Schizosaccharomyces pombe"
                     /strain="972h-"
                     /db_xref="taxon:4896"
                     /chromosome="I"
                     /map="IL"
                     /clone="cosmid c589"
     misc_feature    1. .89
                     /note="nominal overlap with SPAC16A10 S. pombe chromosome 1"
     gene            complement(join(1. .65,121. .777))
                     /gene="SPAC589.01c"
                     /note="SPAC16A10.08c"
     CDS             complement(join(1. .65,121. .777))
                     /partial
                     /gene="SPAC589.01c"
                     /note="SPAC589.01c, very low SIMILARITY:Caenorhabditis elegans., Q21148, k02f3.4 protein., (308 aa), fasta scores: opt: 111, E():2.4, (23.3% identity in 193 aa)"
                     /codon_start=1
                     /label=SPAC589.01c
                     /product="hypothetical protein"
                     /protein_id="CAC19758.1"
                     /db_xref="GI:12043545"
                     /translation="MKKHKSSIKCFKIDQISQQPETDLVFTKRSECKVKDELFSDKEN
                     SILCKQLKELDLVVSSNKEFLNEKTSDQISFLKPRETVVEKKLANGSIWPNETSHLDK
                     SRDLSSHSNGLDALAMTPYIVKENNKTHLSPRGNYPSEAEKSCDFYGQPLHFYRENTS
                     PCLYGSSLSNVFTEKSSDYRHGLGSPTSLVSSNTASKLHLGRKDIVKLSELRNCINSS
                     KDSNQIQELENLLRKEKQRNTE"
     misc_feature    complement(66. .81)
                     /gene="SPAC589.01c"
                     /note="ctgaccaagcgaatag, splice branch and acceptor"
     misc_feature    complement(115. .120)
                     /gene="SPAC589.01c"
                     /note="gtacgt, splice donor sequence"
     gene            complement(join(967. .4482,4518. .4667,4770. .4775))
                     /gene="SPAC589.02c"
     CDS             complement(join(967. .4482,4518. .4667,4770. .4775))
                     /gene="SPAC589.02c"
                     /note="SPAC589.02c, len:1224"
                     /codon_start=1
                     /label=SPAC589.02c
                     /product="hypothetical protein"
                     /protein_id="CAC19759.1"
                     /db_xref="GI:12043546"
                     /translation="MSVCFKKYVKMRQWPHVLSLRVYASYFLLYNNLPLSSAMVHDGV
                     HLISIKTILTDQPCPNTIYYAKYQVQRKDNEDAASLLEDKEAYLRNQDCIVHAKNDLL
                     FVYDFQAIPSIPEESSSFMLLNSGAFSRLALFQKDELALLLDLYINFLQGLKKTVLYW
                     LCKEYNFVPIYGVLLPLKLHPSFDTQLWNIYGYPVVDLQLSVLSKGHIEFYLKPTRQT
                     VYRLSEVDNLVKKLDTVIRLAPTGCLATLTSVHANASAQTVDALKHRYGFSLTTTSKW
                     VGVTLESSALEFSWPLELCFLETSALRMNDDSLSSNLTDLNNLVLPSNVVNNKKELTE
                     FANEEAEASDKRKEGFTEKEETADAVVTLVPSHSSSPVNYSINSAKSTPASIKVNEEI
                     LVADHNVSDDILMEEIDDVGITEADFDYFDLPNVEEKVEMIEPNFANTMTTLDNEEIN
                     TSISQSNTSPNLNTHENIPKQMEIQSDDRMVTEDLNPYNVEVDIPEISLNISDSKIPT
                     SAYMPSYYSAVIFPSSISSIFQKYNYGGKYWCPSPSLSTEDLFESFSVAESVTSTDED
                     ICSTNFIQQDFTMEYNHDFFSSSKTPTNISEQSNPDSNYDTLSLAHQVLMNESKSANF
                     DFSFLKSLDLQPTITLGKNDLLNAILSQNLWFRSLPFWKSMTTSFMMSQDVLNFSSYM
                     RKPIRDYLEKILLGESSAVFLSKSPENYLSSINNGHHALNDNPPSQVNFSETLVNFSQ
                     PPRVLLKYNEKKLSLDSSAPENWISLCLQPYGESKDFEVFLLSSKSPDVSSKAISFFY
                     DVQLAYENCKLGKLNLSETSINERVMGFSTNINETDNYDDNETTQSDTATSYEQLASV
                     CVNELSGKNVLFFYFLEDDSEKLLKACQHFICVKDSIKRLGDNKFEDKSLRICTIPNS
                     IFDSPNSHTTNSNSFFTKVSLDIYNNDPLLMDGSLKRREPAFLLKKPLLSTLNYQLKD
                     INPRSSALGEYALHVTYTTVEEHLLICNWNDSYGEFETERRYFLQDLEIEDALQQILE
                     VTFTFLNSMHMDWIVIVMKIGEMSDAEYLFWDQAIIPENLQGNVSLTVGYCSAEHGPG
                     STSKVFSRIPYSASSTVIRNNSSHELSLVAFIREMAMPVPNDEFKKISTILARGYLAL
                     DEDESYLPLLSIHLLISRNHDPYLMLNLILKHYLSMIYLQFRTYVSFSSLPLHISTVL
                     YQKQLLQFMASDITHPVTS"
     misc_feature    complement(4483. .4502)
                     /gene="SPAC589.02c"
                     /note="ataacatgataaattaaaag, splice branch and acceptor"
     misc_feature    complement(4512. .4517)
                     /gene="SPAC589.02c"
                     /note="gtaatt, splice donor sequence"
     misc_feature    complement(4668. .4686)
```

```
                /gene="SPAC589.02c"
                /note="ttaacagtaaaagtatcag, splice branch and acceptor"
     misc_feature    complement(4764. .4769)
                /gene="SPAC589.02c"
                /note="gtatat, splice donor sequence"
     gene            complement(join(5153. .5311,5375. .5722))
                /gene="SPAC589.03c"
     CDS             complement(join(5153. .5311,5375. .5722))
                /gene="SPAC589.03c"
                /note="SPAC589.03c, len:169, SIMILARITY:Plasmodium
                falciparum., O96229, hypothetical 78.6 kda protein., (665
                aa), fasta scores: opt: 182, E():0.00088, (28.0% identity
                in 161 aa)"
                /codon_start=1
                /label=SPAC589.03c
                /product="hypothetical protein"
                /protein_id="CAC19760.1"
                /db_xref="GI:12043547"
                /translation="MDLSLFSDIKDVVERDMKKNSKVVENSSARSLKNKTEENKTIDK
                HKSSVGPKEPKSSLEKPKANRKRKQSYGLLHDIKELGGTEEDLELIEDVESDEELEFE
                EPNRKQPKKDGTDAFANELNLFAKKLGFSKNSFDARALDTESEDETEIEKSSSENSDS
                VDVDSDDE"
     misc_feature    complement(5312. .5330)
                /gene="SPAC589.03c"
                /note="ctaatatcctctgttttag, splice branch and acceptor"
     misc_feature    complement(5369. .5374)
                /gene="SPAC589.03c"
                /note="gtaagt, splice donor sequence"
     gene            join(5999. .6559,6639. .6893)
                /gene="SPAC589.04"
     CDS             join(5999. .6559,6639. .6893)
                /gene="SPAC589.04"
                /note="SPAC589.04, len:272, SIMILARITY:Homo sapiens,
                O75431, metaxin 2., (263 aa), fasta scores: opt: 220,
                E():2.1e-07, (29.2% identity in 216 aa)"
                /codon_start=1
                /label=SPAC589.04
                /product="mammalian metaxin 2 homologuue; putative
                mitochondrial outer membrane protein putative component of
                a preprotein import complex"
                /protein_id="CAC19761.1"
                /db_xref="GI:12043548"
                /translation="MGGLTKFSSYFHSIFSRFPLITFSNPYPGENEDYKTKTVMYLTM
                WNSDLNSEALDVNSLQWQTWAKLNDPSIVFLNVSNHASPDEKVPFIQIESRKLVLNPS
                LLQYFLKDESTLQQISPWMSLLINQVETAILLTMYLDNENFSEIQKKWLPNMSWPLNI
                IKSIGLPSQIKRKICLQLNESTLDFDAILEDASKAFSALSELLGSDKWFFNNESPSFL
                DVSLFAHAEIINHLPLKNDQLKVVLGTHKNLTDLTTRVRTLAGYTSAGPIALR"
     misc_feature    6560. .6565
                /gene="SPAC589.04"
                /note="gtacgt, splice donor sequence"
     misc_feature    6625. .6638
                /gene="SPAC589.04"
                /note="ctaacaaagcttag, splice branch and acceptor"
     gene            complement(join(7189. .8041,8175. .8323,8370. .8408))
                /gene="SPAC589.05c"
     CDS             complement(join(7189. .8041,8175. .8323,8370. .8408))
                /gene="SPAC589.05c"
                /note="SPAC589.05c, len:347, SIMILARITY:Neurospora
                crassa., CAB98246, conserved hypothetical protein., (372
                aa), fasta scores: opt: 997, E():0, (46.3% identity in 352
                aa)"
                /codon_start=1
                /label=SPAC589.05c
                /product="conserved hypothetical protein; putative
                glycoprotein"
                /protein_id="CAC19762.1"
                /db_xref="GI:12043549"
                /translation="MSRVLQDAEFISLNSNDVKVNKGGCAAVATWIKEKLDSLGPQFA
                EWQNHELHPKTRDVSTLDWIFLVDILNFSFWSDVDVEDSGKHSKRFSIEYKGKLYTGY
                WSLCAAINKALDAGIPITSPAFYADEKQCPDTLIASVFDSATVEKIPLLEERIRIMRA
                SGRVLVDSYHGSYCGLLKKCHNQAQRLIKLLLADFPDFRDVSVYKGRECYMLKRAQIL
                VAETWACFQGQNYGRFDDIDSITMFADYRVPQILWQLGCLSYSSDFKKRLLKNELIAH
                NDPMEIEMRGCSIWAVEKILQNINRKDVNAITIDFFLWDLAKEWQAKGYKPSTQVDEV
                TIPCIRVRSIYY"
     misc_feature    complement(8042. .8059)
                /gene="SPAC589.05c"

Alignment Scores:
Pred. No.:              5.5e-11     Length:          18342
Score:                  264.50      Matches:         103
Percent Similarity:     41.08%      Conservative:    72
Best Local Similarity:  24.18%      Mismatches:      188
Query Match:            7.11%       Indels:          64
DB:                     8           Gaps:            11

US-09-809-638-2 (1-699) x SPAC589 (1-18342)

Qy      11 GluSerLeuLeuGlyCysValSerTrpSerLeuTyrHisAspLeuGlyProMetIleTyr 30
Db   17129 GAGGTGTATCTTTCTTTTGTCTTCTGGAGCGTTCTTACTAGCTTAGGGTTATTAGTATGG 17188

Qy      31 TyrPheProLeuGlnThrLeuGluLeuThrGlyLeuGluGlyPheSerIleAlaPheLeu 50
Db   17189 TATTTTCCCTTATGGCATATGGGGATTTCGGGCTATGAAGCGTGCATTCTTTTCGAACTT 17248

Qy      51 SerProIlePheLeuThrIleThrProPheTrpLysLeuValAsnLysLysTrpMetLeu 70
Db   17249 TCTCCATTCTTACTTGGTATTCCACTATTGCGAAAATTTGCCTCAAAAGTACCTGTTATA 17308

Qy      71 ThrLeuLeuArgIleIleThrIleGlySerIleAlaSerPheGlnAlaProAsnAlaLys 90
Db   17309 TTTTTA------TTTTTAAATGTTATAGGTATAGCTGCCTATAAGCTTGAGGACCCGGTG 17362

Qy      91 LeuArgLeuMetValLeuAlaLeuGlyValSerSerSerLeuIleValGlnAlaValThr 110
Db   17363 CACAGATTATTTGTTACTGCGTTT-----------TCTGTCTGTTGTGAATGTCTCGCC 17410

Qy     111 TrpTrp----------------SerGlySerHisLeuGlnArgTyrLeuArgIleTrp 124
Db   17411 TGGACTTCTTTATTTTCAAATATTTCGCCTGAAAATTTAGCAATAGAACGTAAAATTAGC 17470

Qy     125 GlyPheIleLeuGlyGlnIleValLeuValValLeuArgIleTrpTyrThrSerLeuAsn 144
Db   17471 ACATTTTTGTTTGGATTGCTTGCGTCTTCAATCGCCAAGTACTCTTTTTTTTCTAACAAT 17530

Qy     145 ProIleTrpSerTyrGlnMetSerAsnLysValIleLeuThr---------------- 158
Db   17531 CCTATATGGCC-TATTTTAAATGAAACAAACGGTGGAAAACAAATCCCAGCGTTAATTGT 17589

Qy     159 ----------LeuSerAlaIleAlaThrLeuAspArgIleGlyThrAspGlyAspCys 174
Db   17590 CGGAATAATTGCATGTCTAATATTTGCCATTTTCCATGTGCAGCAAACAACAGCGAATGC 17649

Qy     175 SerLysProGluGluLysLysThrGlyGluValAlaThrGlyMetAlaSerArgProAsn 194
Db   17650 CGT----------------------------GGAACATTTAAACTAAGGAAGATAAC 17679

Qy     195 Trp-LeuLeuAlaGlyAlaAlaPheGlySerLeuValPheLeuThrHisTrpValPheGl 214
Db   17680 TGCTCTGTCTGCAGCCCTAAGCTTGGGCACAGTACTTTTTTGTCTGCATACTTTTCTTTG 17739

Qy     214 yGluValSerLeuValSerArgTrpAlaValSerGlyHisProHisProGlyProAspPr 234
Db   17740 TGACTCTACAGTTCTTATGACATGGTCGTGGGATGGGTATCCTATTAAGGGACCGCAACC 17799

Qy     234 oAsnProPheGlyGlyAlaValLeuLeuCysLeuAlaSerGlyLeuMetLeuProSerCy 254
Db   17800 TTATCCACACGGTGCCGTTTCAATAGTTGTTTCCATTTGCGCTGTTTTGGTAGCCCCTTA 17859

Qy     254 sLeuTrpPheArgGlyThrGlyLeuIleTrpTrpValThrGlyThrAlaSerAlaAlaGl 274
Db   17860 TTTATATCAATCAGGTGCTTTTATGCTTATTGGATTTGTTTTAGCCTGCTTCGGATCATA 17919

Qy     274 yLeuLeuTyrLeu---HisThrTrpAlaAlaAlaValSerGlyCysValPheAlaIlePh 293
Db   17920 TTTTATGTATATTAACCATGGATGGTGTAGTTATTTGGGTGGGTTAATATTTACTTCCTA 17979
```

```
Qy     293 eThr---------------AlaSerMetTrpProGlnThrLeuGlyHisLeuIleAsnSe 308
           :                 ||||||:::                         |||:::||
Db   17980 CGTTTTAATTTACTCGTTCGCTTCTATC--------------------AGAATTTCCTC 18018

Qy     308 rGlyThrAsnProGlyLysThrMetThrIleAlaMetIlePheTyrLeuLeuGluIlePh 328
           |      :::|||  |||            |||   :::   |||:::|||
Db   18019 CTTCTATTCTCCTGCGAAGGTTTGGGGTGGCGCTTTTTTGGTTTACATCCTTTACTCATT 18078

Qy     328 ePheCysAlaTrpCysThrAlaPheLysPheValProGlyGlyValTyrAlaArgGluAr 348
           |||      |||::::::|||||||||||||||         ||||||||
Db   18079 AGCTCATGTTTGGGTCGTAGCCTACGAGTTTGTTCCAGGTGGACCCATTCTTCGAGAGAG 18138

Qy     348 gSerAspValLeuLeuGlyThrMetMetLeuIleIleGlyLeuAsnMetLeuPhe----- 366
           |:::      :::|||            :::   ||||||   |||:::
Db   18139 AACTTCTTACATTCTA-----------ATTTTCATTGGTTGGAACCTTGCAGCTTTAGT 18186

Qy     367 ------------------------GlyProLysLysAsnLeuAspLeuLeuLeuGlnTh 378
                                      |||   |||           ::::::
Db   18187 ACCTGCGTATTCGGGTGAGTCCAAAGAACCCAACAAAGCAGACTCATCTGTAGTAGATAT 18246

Qy     378 rLysAsnSerSerLysValLeuPheArgLysSer-GluLys------TyrMetLysLeuP 396
           |||   |||           |||:::|||    |||      |||   ::::::|
Db   18247 CAAACAATCAGATTCTTCTTATCGTAGAAGAAGTTTCAAAAAATCACTATTGACAGGTTT 18306

Qy     396 heLeuTrpLeuLeu 400
           ||:::||||||||
Db   18307 TTGTCTGGCTCTTA 18320
```

```
RESULT 11
AC106576/c
LOCUS       AC106576              165893 bp    DNA     linear   HTG 12-JAN-2002
DEFINITION  Rattus norvegicus clone CH230-198N10, *** SEQUENCING IN PROGRESS
            ***, 68 unordered pieces.
ACCESSION   AC106576
VERSION     AC106576.1  GI:18139100
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 165893)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J.,
            Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
            Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
            Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
            Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
            Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
            Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
            Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
            Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
            Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
            Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
            Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
            Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
            Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
            Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
            Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
            Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
            Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
            Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
            Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
            Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
            Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
            Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
            Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
            Morgan,M., Morris,S., Moser,M., Neal,D., Newtson,J., Newtson,N.,
            Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
            Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
            Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
            Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M.,
            Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N.,
            Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H.,
            Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
            Tang,H., Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N.,
            Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R.,
            Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
            Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S.,
            Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 165893)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-JAN-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     -------------- Genome Center
              Center: Baylor College of Medicine
              Center code: BCM
              Web site: http://www.hgsc.bcm.tmc.edu/
              Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
              Center project name: GLJY
              Center clone name: CH230-198N10
            -------------- Summary Statistics
              Assembly program: Phrap; version 0.990329First call to
            findPhrapList
              Consensus quality: 137306 bases at least Q40
              Consensus quality: 144043 bases at least Q30
              Consensus quality: 149677 bases at least Q20
              Estimated insert size: 139846; sum-of-contigs estimation
              Quality coverage: 0x in Q20 bases; agarose-fp estimation
              Quality coverage: 2.2x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 68 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     7956: contig of 7956 bp in length
            *     7957     8056: gap of unknown length
            *     8057    13903: contig of 5847 bp in length
            *    13904    14003: gap of unknown length
            *    14004    17958: contig of 3955 bp in length
            *    17959    18058: gap of unknown length
            *    18059    22354: contig of 4296 bp in length
            *    22355    22454: gap of unknown length
            *    22455    26589: contig of 4135 bp in length
            *    26590    26689: gap of unknown length
            *    26690    31248: contig of 4559 bp in length
            *    31249    31348: gap of unknown length
            *    31349    36556: contig of 5208 bp in length
            *    36557    36656: gap of unknown length
            *    36657    40383: contig of 3727 bp in length
            *    40384    40483: gap of unknown length
            *    40484    44415: contig of 3932 bp in length
            *    44416    44515: gap of unknown length
            *    44516    49038: contig of 4523 bp in length
            *    49039    49138: gap of unknown length
            *    49139    52349: contig of 3211 bp in length
            *    52350    52449: gap of unknown length
            *    52450    56176: contig of 3727 bp in length
            *    56177    56276: gap of unknown length
            *    56277    60332: contig of 4056 bp in length
            *    60333    60432: gap of unknown length
            *    60433    62803: contig of 2371 bp in length
            *    62804    62903: gap of unknown length
```

```
*    62904    66621: contig of 3718 bp in length
*    66622    66721: gap of unknown length
*    66722    69020: contig of 2299 bp in length
*    69021    69120: gap of unknown length
*    69121    72220: contig of 3100 bp in length
*    72221    72320: gap of unknown length
*    72321    75607: contig of 3287 bp in length
*    75608    75707: gap of unknown length
*    75708    78960: contig of 3253 bp in length
*    78961    79060: gap of unknown length
*    79061    81780: contig of 2720 bp in length
*    81781    81880: gap of unknown length
*    81881    85140: contig of 3260 bp in length
*    85141    85240: gap of unknown length
*    85241    87883: contig of 2643 bp in length
*    87884    87983: gap of unknown length
*    87984    90762: contig of 2779 bp in length
*    90763    90862: gap of unknown length
*    90863    93086: contig of 2224 bp in length
*    93087    93186: gap of unknown length
*    93187    96030: contig of 2844 bp in length
*    96031    96130: gap of unknown length
*    96131    98070: contig of 1940 bp in length
*    98071    98170: gap of unknown length
*    98171   100910: contig of 2740 bp in length
*   100911   101010: gap of unknown length
*   101011   102968: contig of 1958 bp in length
*   102969   103068: gap of unknown length
*   103069   104748: contig of 1680 bp in length
*   104749   104848: gap of unknown length
*   104849   106306: contig of 1458 bp in length
*   106307   106406: gap of unknown length
*   106407   108402: contig of 1996 bp in length
*   108403   108502: gap of unknown length
*   108503   110234: contig of 1732 bp in length
*   110235   110334: gap of unknown length
*   110335   112075: contig of 1741 bp in length
*   112076   112175: gap of unknown length
*   112176   113608: contig of 1433 bp in length
*   113609   113708: gap of unknown length
*   113709   115299: contig of 1591 bp in length
*   115300   115399: gap of unknown length
*   115400   116670: contig of 1271 bp in length
*   116671   116770: gap of unknown length
*   116771   119076: contig of 2306 bp in length
*   119077   119176: gap of unknown length
*   119177   120752: contig of 1576 bp in length
*   120753   120852: gap of unknown length
*   120853   122336: contig of 1484 bp in length
*   122337   122436: gap of unknown length
*   122437   125169: contig of 2733 bp in length
*   125170   125269: gap of unknown length
*   125270   126815: contig of 1546 bp in length
*   126816   126915: gap of unknown length
*   126916   128187: contig of 1272 bp in length
*   128188   128287: gap of unknown length
*   128288   129839: contig of 1552 bp in length
*   129840   129939: gap of unknown length
*   129940   131660: contig of 1721 bp in length
*   131661   131760: gap of unknown length
*   131761   133894: contig of 2134 bp in length
*   133895   133994: gap of unknown length
*   133995   135658: contig of 1664 bp in length
*   135659   135758: gap of unknown length
*   135759   136862: contig of 1104 bp in length
*   136863   136962: gap of unknown length
*   136963   138366: contig of 1404 bp in length
*   138367   138466: gap of unknown length
*   138467   139604: contig of 1138 bp in length
*   139605   139704: gap of unknown length
*   139705   141083: contig of 1379 bp in length
*   141084   141183: gap of unknown length
*   141184   142432: contig of 1249 bp in length
*   142433   142532: gap of unknown length
*   142533   144347: contig of 1815 bp in length
*   144348   144447: gap of unknown length
*   144448   145945: contig of 1498 bp in length
*   145946   146045: gap of unknown length
*   146046   147800: contig of 1755 bp in length
*   147801   147900: gap of unknown length
*   147901   149217: contig of 1317 bp in length
*   149218   149317: gap of unknown length
*   149318   150337: contig of 1020 bp in length
*   150338   150437: gap of unknown length
*   150438   151919: contig of 1482 bp in length

Alignment Scores:
Pred. No.:              0.0053          Length:         165893
Score:                  182.50          Matches:        39
Percent Similarity:     89.36%          Conservative:   3
Best Local Similarity:  82.98%          Mismatches:     5
Query Match:            4.91%           Indels:         1
DB:                     2               Gaps:           0

US-09-809-638-2 (1-699) x AC106576 (1-165893)

Qy    628 IleSerHisAlaGluLeuSerAspSerGluIleGlnMetAlaLysPheArgIleProAsp 647
          ||||||||||||||||||||||||||||||||:::  ||||||||||||||||||||
Db 165892 ATTTCTCATGCTGAACTGAGTGACTCTGAGATCAGAT--GCCAAATTCAGGATCCCAGAT 165835

Qy    648 AspProThrAsnTyrArgAspAsnGlnLysValValIleAspHisArgGluValSerGlu 667
          ||||||  ||||||||||||||||:::||||||||||||||||  |||   :::
Db 165834 GATCCCTTCAATTACAGAGACAACCAGAGAGTCGTCATAGACCACAGAGGAGTTCCCAAG 165775

Qy    668 LysIleHisPheAsnProArg 674
          ||||||||||||||||||
Db 165774 AACATTCATTTCAACCCGAGG 165754

RESULT 12
AX337478/c
LOCUS       AX337478               305 bp    DNA     linear   PAT 09-JAN-2002
DEFINITION  Sequence 7987 from Patent WO0194629.
ACCESSION   AX337478
VERSION     AX337478.1  GI:18128197
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Young,P.E., Augustus,M., Carter,K.C., Ebner,R., Endress,G.,
            Horrigan,S., Soppet,D.R. and Weaver,Z.
  TITLE     Cancer gene determination and therapeutic screening using signature
            gene sets
  JOURNAL   Patent: WO 0194629-A 7987 13-DEC-2001;
            Avalon Pharmaceuticals (US)
FEATURES             Location/Qualifiers
     source          1. .305
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT       81 a     45 c     44 g    127 t      8 others
ORIGIN

Alignment Scores:
Pred. No.:              1.38e-06        Length:         305
Score:                  182.00          Matches:        31
Percent Similarity:     91.18%          Conservative:   0
Best Local Similarity:  91.18%          Mismatches:     3
Query Match:            4.89%           Indels:         0
DB:                     6               Gaps:           0

US-09-809-638-2 (1-699) x AX337478 (1-305)

Qy    664 GluValSerGluLysIleHisPheAsnProArgPheGlySerTyrLysGluGlyHisAsn 683
          |||||||||  ||||||||||||||||||  ||||||||||||||||||||||||||||
Db    300 GAAGTTTCTGNGAAAATTCATTTTAATCCCNGATTTGGATCCTACAAAGAAGGNCACAAT 241
```

```
Qy     684 TyrGluAsnAsnHisHisPheHisMetAsnThrProLysTyr 697
           ||||||||||||||||||   ||||||||||||||||||||
Db     240 TATGAAAACAACCATCATTNTCATATGAATACTCCCAAATAC 199

RESULT 13
CNS06EJ2
LOCUS       CNS06EJ2                849 bp    DNA     linear   STS 10-JAN-2001
DEFINITION  T3 end of clone AR0AA022A01 of library AR0AA from strain CBS 732 of
            Zygosaccharomyces rouxii, sequence tagged site.
ACCESSION   AL395172
VERSION     AL395172.1  GI:12146608
KEYWORDS    STS.
SOURCE      Zygosaccharomyces rouxii.
  ORGANISM  Zygosaccharomyces rouxii
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Saccharomycetaceae; Zygosaccharomyces.
REFERENCE   1  (bases 1 to 849)
  AUTHORS   de Montigny,J., Straub,M., Potier,S., Tekaia,F., Dujon,B.,
            Wincker,P., Artiguenave,F. and Souciet,J.
  TITLE     Genomic Exploration of the Hemiascomycetous Yeasts: 8.
            Zygosaccharomyces rouxii(1)
  JOURNAL   FEBS Lett. 487 (1), 52-55 (2000)
  PUBMED    11152883
REFERENCE   2  (bases 1 to 849)
  AUTHORS   Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
            Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,
            de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B.,
            Malpertuy,A., Neuveglise,C., Ozier-Kalogeropoulos,O., Potier,S.,
            Saurin,W., Tekaia,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,
            Wincker,P. and Weissenbach,J.
  TITLE     Genomic Exploration of the Hemiascomycetous Yeasts: 1. A set of
            yeast species for molecular evolution studies(1)
  JOURNAL   FEBS Lett. 487 (1), 3-12.(2000)
  PUBMED    11152876
REFERENCE   3  (bases 1 to 849)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-SEP-2000) Genoscope - Centre National de Sequencage,
            2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
            seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT     This STS is part of a random genomic sequencing program of thirteen
            yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
            exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
            Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
            lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
            angusta, Debaryomyces hansenii var.  hansenii, Pichia sorbitophila,
            Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
            5 kb were prepared and both extremities were sequenced. See
            keywords for description of this sequence and for the sequence of
            the other extremity of this insert.
FEATURES            Location/Qualifiers
     source         1. .849
                    /organism="Zygosaccharomyces rouxii"
                    /strain="CBS 732"
                    /db_xref="taxon:4956"
                    /clone="AR0AA022A01"
                    /clone_lib="AR0AA"
                    /note="end : T3"
     misc_feature   <5. .>847
                    /note="similar to Saccharomyces cerevisiae ORF YCR017c [
                    similarity to hypothetical S.pombe protein ]"
                    /evidence=not_experimental
BASE COUNT      198 a    157 c    193 g    300 t      1 others
ORIGIN

Alignment Scores:
Pred. No.:              0.0014          Length:         849
Score:                  152.50          Matches:        71
Percent Similarity:     33.23%          Conservative:   34
Best Local Similarity:  22.47%          Mismatches:     120
Query Match:            4.10%           Indels:         91
DB:                     11              Gaps:           13
```

```
US-09-809-638-2 (1-699) x CNS06EJ2 (1-849)

Qy      87 ProAsnAlaLysLeuArgLeuMetValLeuAlaLeuGlyValSerSerSerLeuIleVal 106
           |||   ::::::|||           |||      |||         :::|||:::|||
Db      17 CCGGAAAGTCGCCTG-----------CTGGTGACTGGTGCCGGTGATGCTCTTGTGGTG 64

Qy     107 GlnAla---------------------------------------------------------- 108
           |||
Db      65 AGTGCATTTGTACTCAATTTGAAATCTATTAAAAGACCTGAAGTCAATGCCGCTTTCACT 124

Qy     109 ValThrTrpTrpSerGlySerHisLeuGlnArgTyrLeuArgIleTrpGlyPheIleLeu 128
           ||||||||                                              :::|||
Db     125 GTGACATGG-----------------------------------------------TTGTTG 139

Qy     129 GlyGlnIleValLeuValValLeuArgIleTrpTyrThrSerLeuAsnProIleTrp--- 147
           |||   ::::::   |||:::|||:::                    ||||||:::|||
Db     140 GGTCTAGTCATCTCTGTGATATTAAAATTCGCCTGCCATGGTAACAATCCCTTGTGGCCT 199

Qy     148 ----------------------SerTyrGlnMetSerAsnLysValIleLeuThrLeuSer 160
                                  |||                :::      |||:::
Db     200 ATTACTAAAGCTGAAAATGGTGGCTATCACAACATTGGATTGATTTTCACAACCATTTTT 259

Qy     161 AlaIleAlaThrLeuAspArgIleGlyThrAspGlyAspCysSerLysProGluGluLys 180
           |||              :::   :::                |||        |||
Db     260 GGAATCTTGGCA---CCTTATGTCAATTCATTACATTTTGCTCCCAAAGATAGTTCTAAG 316

Qy     181 LysThrGlyGluValAlaThrGlyMetAlaSerArgProAsnTrpLeuLeuAlaGlyAla 200
           |||:::            :::  |||                          |||
Db     317 GCTGACGGAAAGACTACGGAATCTATAAAGAGTAACAAATACAAGTTTTTTGTGGGTCTT 376

Qy     201 AlaPheGlySerLeuValPheLeuThrHisTrpValPheGlyGluValSerLeuValSer 220
           ||||||||:::::::|||      |||      |||   :::   |||   :::
Db     377 GGATTCGGTTCCATAATTTTTAGTCTTCACCGATTCTTTACCGATGCCTCTGCATTGATT 436

Qy     221 ArgTrpAlaValSerGly------HisProHisProGlyProAspProAsnProPheGly 238
           |||:::     |||                  ||||||        |||:::
Db     437 TATTGGTCTTGGGAGGGTTGGAACCAAGAAACTCAAGGCCCCTTGGCTTGGCCATGGAGT 496

Qy     239 GlyAlaValLeuLeuCysLeuAlaSerGlyLeu---MetLeuProSerCysLeuTrpPhe 257
           |||      |||   ||||||          |||      |||   |||      :::
Db     497 GGG------TTAACTTGCCTTGTCATGCTATTAGCTTCTTTAACCTCCGTGAGATTCTCT 550

Qy     258 ArgGlyThrGlyLeuIleTrpTrpValThrGlyThrAlaSerAlaAlaGlyLeuLeuTyr 277
           :::   |||            ||||||   |||                  ||||||
Db     551 AAAAATACA-----------TGGGTACCTGGTATC--------------CTTTGTTG 583

Qy     278 LeuHisThrTrpAlaAlaAla------------------------------------- 284
           ||||||
Db     584 GAATCTACTTGGGTCTTGTGCAATAGAAACATTACTGAATGGGATAACTATCTTTACGGT 643

Qy     285 ------------ValSerGlyCysValPheAlaIle---------PheThrAlaSerMet 297
                        :::   |||   ::::::   |||         |||   |||
Db     644 GGGTTACCTTATATCTTGGGTGTGATTTGGACCATTCCATCATATTTTGCAGCTTTGAGT 703

Qy     298 TrpProGlnThrLeuGlyHisLeuIleAsnSerGlyThrAsnProGlyLysThrMetThr 317
           |||:::::::|||   :::|||
Db     704 GAATTACAGAGTGTTGGCGTCATAATT-------------------------------- 730

Qy     318 IleAlaMetIlePheTyrLeuLeuGluIlePhePheCysAlaTrpCysThrAlaPheLys 337
           :::   :::   |||          |||      |||      |||:::
Db     731 CTAGGTTTTGTTACCTATGCGTTCATCAGTTTTGCCACTGTTTGGACCGTTGCGTATGCA 790

Qy     338 PheValProGlyGlyValTyrAlaArgGluArgSerAspValLeuLeu 353
           ||||||||||   |||          ||||||   |||   :::   |||
Db     791 TTTGTTCCTCTTGGATGGTTATTGAGAGAAGAATCGAGGTTGTATTTG 838

RESULT 14
AC096364/c
LOCUS       AC096364              225532 bp    DNA     linear   HTG 20-DEC-2001
DEFINITION  Rattus norvegicus chromosome Rf1 clone CH230-59C15, WORKING DRAFT
            SEQUENCE, 20 unordered pieces.
ACCESSION   AC096364
```

```
VERSION     AC096364.4  GI:17944066
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 225532)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J.,
            Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
            Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
            Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
            Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
            Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
            Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
            Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
            Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
            Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
            Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
            Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
            Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
            Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
            Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
            Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
            Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
            Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
            Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
            Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
            Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
            Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
            Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
            Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
            Morgan,M., Morris,S., Moser,M., Neal,D., Newtson,J., Newtson,N.,
            Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
            Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
            Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
            Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M.,
            Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N.,
            Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H.,
            Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
            Tang,H., Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N.,
            Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R.,
            Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
            Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S.,
            Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 225532)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-SEP-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Dec 20, 2001 this sequence version replaced gi:15920027.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GEZQ
                Center clone name: CH230-59C15
            -------------- Summary Statistics
                Assembly program: Phrap; version 0.990329First call to
            findPhrapList
                Consensus quality: 220440 bases at least Q40
                Consensus quality: 221708 bases at least Q30
                Consensus quality: 222752 bases at least Q20
                Estimated insert size: 216271; sum-of-contigs estimation
                Quality coverage: 0x in Q20 bases; agarose-fp estimation
                Quality coverage: 10.2x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 20 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1    37343: contig of 37343 bp in length
            *   37344    37443: gap of unknown length
            *   37444    59505: contig of 22062 bp in length
            *   59506    59605: gap of unknown length
            *   59606    96908: contig of 37303 bp in length
            *   96909    97008: gap of unknown length
            *   97009   126921: contig of 29913 bp in length
            *  126922   127021: gap of unknown length
            *  127022   144976: contig of 17955 bp in length
            *  144977   145076: gap of unknown length
            *  145077   163948: contig of 18872 bp in length
            *  163949   164048: gap of unknown length
            *  164049   170970: contig of 6922 bp in length
            *  170971   171070: gap of unknown length
            *  171071   179611: contig of 8541 bp in length
            *  179612   179711: gap of unknown length
            *  179712   186544: contig of 6833 bp in length
            *  186545   186644: gap of unknown length
            *  186645   192903: contig of 6259 bp in length
            *  192904   193003: gap of unknown length
            *  193004   201690: contig of 8687 bp in length
            *  201691   201790: gap of unknown length
            *  201791   206903: contig of 5113 bp in length
            *  206904   207003: gap of unknown length
            *  207004   209367: contig of 2364 bp in length
            *  209368   209467: gap of unknown length
            *  209468   214422: contig of 4955 bp in length
            *  214423   214522: gap of unknown length
            *  214523   216975: contig of 2453 bp in length
            *  216976   217075: gap of unknown length
            *  217076   219621: contig of 2546 bp in length
            *  219622   219721: gap of unknown length
            *  219722   221777: contig of 2056 bp in length
            *  221778   221877: gap of unknown length
            *  221878   222899: contig of 1022 bp in length
            *  222900   222999: gap of unknown length
            *  223000   224359: contig of 1360 bp in length
            *  224360   224459: gap of unknown length
            *  224460   225532: contig of 1073 bp in length.
FEATURES             Location/Qualifiers
     source          1. .225532
                     /organism="Rattus norvegicus"
                     /db_xref="taxon:10116"
                     /chromosome="Rf1"
                     /clone="CH230-59C15"
BASE COUNT    61510 a   52080 c   49274 g   60665 t    2003 others
ORIGIN

Alignment Scores:
Pred. No.:                10.3          Length:         225532
Score:                    144.50        Matches:        156
Percent Similarity:       36.41%        Conservative:   77
Best Local Similarity:    24.38%        Mismatches:     213
Query Match:              3.88%         Indels:         198
DB:                       2             Gaps:           33

US-09-809-638-2 (1-699) x AC096364 (1-225532)

Qy     100 ValSerSerSerLeuIleValGlnAlaValThrTrpTrpSerGlySerHis--------- 116
               |||   :::|||||||:::      |||::::::   ||||||||||
Db 214157 GTCCACACATCTTTACTT------GCAATGAGTCATTGGTCTGGTTCGAGGTCTCTGGCT 214104
```

```
Qy    117 ------------------LeuGlnArgTyrLeuArgIleTrpGlyPheIleLeuGlyGln 130
Db 214103 CCTGCTACGTTGTCAATATTGGATCCTCTTGGGGACTTCTCAGATATCCTATTGTTGCCC 214044

Qy    131 IleValLeuValValLeuArgIleTrpTyrThrSerLeuAsnProIleTrpSerTyrGln 150
Db 214043 TGTGTTATGGAGGTCCTACAGCTTTGGATCTGCAGGACCTGCCCCTTCATGAGCTCCAGC 213984

Qy    151 MetSerAsnLysValIleLeuThrLeuSerAlaIleAlaThrLeuAspArgIleGlyThr 170
Db 213983 AGATCAGATGGGGTAGATGTTGGGGTGGGCCAAATCAGAGCCCTG---GTTCTGGGCCTG 213927

Qy    171 AspGlyAsp--------------CysSerLysProGluGluLysLysThrGlyGluVal 185
Db 213926 GGTGGTAGCCGAGCTAGTCAGCCGGCCAGCTCTCCTGCACCCACAGCAGCAGGAGAGCCC 213867

Qy    186 Ala---ThrGlyMetAlaSerArgProAsnTrpLeuLeuAlaGlyAlaAlaPheGlySer 204
Db 213866 TCCGGCACCGCCCTCCCTAACTCACCCAAT---------GCAGCAGCCAGCAAGGGG--- 213819

Qy    205 LeuValPheLeuThrHisTrpValPheGlyGluValSer------LeuValSerArg-Tr 222
Db 213818 ----------------------CTGGGGCAGCTCTCTGCTCTCGTGGCTTCAAGACTG 213783

Qy    222 pAlaValSerGlyHisProHisProGlyProAspProAsnProPheGlyGlyAlaValLe 242
Db 213782 GCTCGCCCAAAACCACCCCCACCCCAACCCACC-CCAGACCCTCTAAACCAGAGCCAGCT 213724

Qy    242 uLeuCysLeuAlaSerGly---------LeuMetLeuProSerCysLeuTrpPheArgGl 259
Db 213723 ATTCTGTGCTGCCCAGGCAAGGTACAGGGCCCTCTCTGTGGAGTGTTGCAGCCCTTGAGG 213664

Qy    259 yThrGlyLeuIleTrpTrpValThr---------GlyThrAlaSerAlaAlaGlyLeuLe 276
Db 213663 GACAGGGACAGCTCTCTGGCTTTCATTGCCTTGGGGCCAGCTCTCCCACCTGCCATATGT 213604

Qy    276 uTyrLeuHisThrTrpAlaAlaAlaValSerGlyCysValPheAlaIlePheThrAlaSe 296
Db 213603 GGTGAGGGACTGGTGG--------------GGGTGACATTTCTCCCTCTTCTATGCCAC 213559

Qy    296 rMetTrpPro-GlnThrLeuGlyHis----------------------------LeuI 306
Db 213558 TGCATGGCAGACAAATAGCAGGGTCAGCTCTCCCATGTTCACATCCTCGGGATTGACTCA 213499

Qy    306 leAsnSerGlyThrAsnProGlyLysThrMetThrIleAlaMetIlePheTyrLeuLeuG 326
Db 213498 TCTGCAACTCCCCAAGCCAGGGACAG---------------------------------- 213473

Qy    326 luIlePhePheCysAla---------------------------------------- 331
Db 213472 -----TTCTACTGTGCTTCCTGTACAAGGTACTGTCCTGGTCTCATGCCTTTGGGGCCAG 213418

Qy    332 ---------------------------------TrpCysThrAlaPheLysPheValP 340
Db 213417 CTCTCCCATGTTCCCCAGGCAAGGGGTGAAGCCAGTTCTGCACAGCCCTCAAAAGTCAAC 213358

Qy    340 roGlyGlyValTyrAlaArgGluArgSerAspValLeuLeu---------GlyThrMetM 357
Db 213357 ATGGCCCCCAAACCACAGCCCAGACCAGGGACGTCTGCCTGGCCTTTAGTGGTAGCTGAC 213298

Qy    357 etLeuIleIleGlyLeuAsnMetLeuPheGlyProLysLysAsnLeuAspLeuLeuLeuG 377
Db 213297 TCCTGCTATG-GCCCTA------------GGAGGCAACAAGGGCCAGGACCTCACCATGG 213251

Qy    377 ln------------ThrLysAsnSerSerLysValLeu-----------------PheA 387
Db 213250 CCTGAGGTAGCATCACCAGATACTCACATCAGGCCCTTCCTCACTACCCTGGAGTCTCCA 213191

Qy    387 rgLysSerGluLysTyr---------------MetLysLeuPheLeu------------ 397
Db 213190 GTTCTGCCTCTCTTTATTGTGCCCCCATCCTTCTGTTTCTCTTTCTCATCTATTCCACCA 213131

Qy    398 -----TrpLeuLeuValGlyValGlyLeuGluGlyLeuGlyLeu---------------- 410
```

```
Db 213130 CTTACTGGCTCATCTCAGCCCAGGGCCTATGAGTGTCTGGAGTCATCTCAGGAGCGGTCT 213071

Qy    411 ---------------------------------ArgHisLysAlaTyr------- 415
Db 213070 CAGAAGAGCTCCACCTGTGCCATGGGACACTGAGCTGGCGCCATCTTGAGTATACTCCGC 213011

Qy    416 --------------------------GluArgLysLeuGlyLysValAlaProThrLysG 427
Db 213010 CCACCTGGGCCTGGGGGCACTGGGCTGAGGGTTGTCTTGGGTATGATTTGCCCACCCAG- 212952

Qy    427 luValSerAlaAlaIleTrp-ProPheArgPheGlyTyrAspAsnGluGlyTrpSerSer 446
Db 212951 ----------GCCACATGGCCCT--------------------GAGTGCTGGTGATCT 212924

Qy    447 LeuGluArgSerAlaHisLeu---LeuAsnGluThrGlyAlaAspPheIleThrIleLeu 465
Db 212923 TGGGAGTTCTCTGCCCACCTGGGCTTGCATGGCACTGGGGTCAGGGTCCTCTTGGAACTG 212864

Qy    466 GluSerAspAlaSerLysProTyrMetGlyAsnAsnAspLeuThrMetTrpLeu---Gly 484
Db 212863 GACTCCAT-GGTGCTAGACTTGGGGTTTTCTCAGGCTTGACCCCTGTCTGGCTGTGTGGT 212805

Qy    485 GluLysLeuGly---PheTyrThrAspPheGlyProSerThrArgTyrHisThrTrpGly 503
Db 212804 CCTGGCTGGGGGTTGTTTTGGGTACGGTTT--------------------GCCTGGGGC 212766

Qy    504 IleMetAlaLeuSerArgTyrProIle---------------ValLysSerGluHisHis 518
Db 212765 CTCATGGCACCAGAAAGGTTTCAAATAAAATAATTTTTAACAGTGCGTGTCCAACCTGTA 212706

Qy    519 LeuLeuProSerProGluGlyGluIleAlaProAlaIleThrLeuThrValAsnIleSer 538
Db 212705 AGATTTCCATC-CCAGAGTATGATGTGATGACTCTCATTTTTATCTCTACCTCACTGTCT 212647

Qy    539 GlyLysLeuValAspPheValValThrHisPheGlyAsnHisGluAspAsp-LeuAspAr 558
Db 212646 ACCCTTCCTCTTTTTATCCACCCTTTCCACTTTGGTTTAGGTCACCTGGATTCTTGGCAG 212587

Qy    558 gLysLeuGlnAla------------IleAlaValSerLysLeuLeuLysSerSerSerAs 574
Db 212586 CAAGTTAAATCTTCCACCTCTATAGATGGCTGAGAGTCCTCTTATC-------------- 212541

Qy    574 nGlnValIlePheLeuGlyTyr-----------IleThrSerAlaProGlySerArgAs 590
Db 212540 ----TAAGTCCCCCTTGGTCATTCCCAGATTCCCTTATTCTGATCCCCAGGTTCAGAAAG 212485

Qy    590 pTyrLeuGlnLeuThrGluHisGlyAsn---ValLysAspIleAspSerThr 606
Db 212484 GTTTGTGATTCTCTCCAGGAAGCAGAACTCCATACAGGATGTGCCTGTCACT 212433
```

```
RESULT 15
SCJ21/c
LOCUS       SCJ21                31717 bp    DNA     linear   BCT 05-AUG-1999
DEFINITION  Streptomyces coelicolor cosmid J21.
ACCESSION   AL109747
VERSION     AL109747.1  GI:5708233
KEYWORDS    ABC transport system ATP-binding protein; bldA codon;
            flavin-containing monooxygenase; oxidoreductase; transcriptional
            regulator multi-domain beta keto-acyl synthase; TTA Leu codon.
SOURCE      Streptomyces coelicolor A3(2).
  ORGANISM  Streptomyces coelicolor A3(2)
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 31717)
  AUTHORS   Redenbach,M., Kieser,H.M., Denapaite,D., Eichner,A., Cullum,J.,
            Kinashi,H. and Hopwood,D.A.
  TITLE     A set of ordered cosmids and a detailed genetic and physical map
            for the 8 Mb Streptomyces coelicolor A3(2) chromosome
  JOURNAL   Mol. Microbiol. 21 (1), 77-96 (1996)
  MEDLINE   97000351
REFERENCE   2  (bases 1 to 31717)
  AUTHORS   Seeger,K.J. and Harris,D.
```

```
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 31717)
  AUTHORS   Parkhill,J., Barrell,B.G. and Rajandream,M.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-AUG-1999) Streptomyces coelicolor sequencing project,
            Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
            CB10 1SA E-mail: barrell@sanger.ac.uk Cosmids supplied by Prof.
            David A. Hopwood, [3] John Innes Centre, Norwich Research Park,
            Colney, Norwich, Norfolk NR4 7UH, UK
COMMENT     Notes:
            Streptomyces coelicolor sequencing at The Sanger Centre is funded
            by the BBSRC and Beowulf Genomics
            Details of S. coelicolor sequencing at the Sanger Centre are
            available on the World Wide Web.
            (URL; http://www.sanger.ac.uk/Projects/S_coelicolor/)
            CDS are numbered using the following system eg SC7B7.01c. SC (S.
            coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary
            strand).
            The more significant matches with motifs in the PROSITE database
            are also included but some of these may be fortuitous.
            The length in codons is given for each CDS.
            Usually the highest scoring match found by fasta -o is given for
            CDS which show significant similarity to other CDS in the database.
            The position of possible ribosome binding site sequences are given
            where these have been used to deduce the initiation codon.
            Gene prediction is based on positional base preference in codons
            using a specially developed Hidden Markov Model (Krogh et al.,
            Nucleic Acids Research, 22(22):4768-4778(1994)) and the FramePlot
            program of Bibb et al., Gene 30:157-66(1984) as implemented at
            http://www.nih.go.jp/
            jun/cgi-bin/frameplot.pl. CAUTION:  We may not have predicted the
            correct initiation codon.  Where possible we choose an initiation
            codon (atg, gtg, ttg or (att)) which is preceded by an upstream
            ribosome binding site sequence (optimally 5-13bp before the
            initiation codon).  If this cannot be identified we choose the most
            upstream initiation codon.
            IMPORTANT: This sequence MAY NOT be the entire insert of the
            sequenced clone.  It may be shorter because we only sequence
            overlapping sections once, or longer, because we arrange for a
            small overlap between neighbouring submissions.
            Cosmid J21 lies between J11 and J33 on the AseI-J genomic
            restriction fragment.
FEATURES             Location/Qualifiers
     source          1. .31717
                     /organism="Streptomyces coelicolor A3(2)"
                     /strain="A3(2)"
                     /db_xref="taxon:100226"
                     /clone="cosmid J21"
     gene            complement(1. .342)
                     /gene="SCJ21.01c"
     misc_feature    1. .136
                     /note="nominal overlap with StJ11"
     CDS             complement(<1. .342)
                     /gene="SCJ21.01c"
                     /note="SCJ21.01c, probable integral membrane protein,
                     partial CDS, len: >114 aa"
                     /codon_start=1
                     /transl_table=11
                     /product="putative integral membrane protein"
                     /protein_id="CAB52347.1"
                     /db_xref="GI:5708234"
                     /translation="MRRHLSLFVTASRYALIGHARNRFAMLLVVVYIPVWIALAYVTI
                     PDQPAPFRLRATGEVLSPPGNHLTQITGALNAVTLIAGFMMFAATFTGGAFDRRLAMA
                     GYPRFHLVLAKL"
     gene            complement(339. .953)
                     /gene="SCJ21.02c"
     CDS             complement(339. .953)
                     /gene="SCJ21.02c"
                     /note="SCJ21.02c, probable ABC transport system
                     ATP-binding protein, len: 204 aa; similar to many e.g.
                     NUC_ECOLI high-affinity zinc uptake system ATP-binding
                     protein (251 aa), fasta scores; opt:  327 z-score: 404.3
                     E(): 3.6e-15, 35.2% identity in 179 aa overlap. Contains
                     PS00017 ATP/GTP-binding site motif A (P-loop) and Pfam
                     match to entry PF00005 ABC_tran, ABC transporter"
                     /codon_start=1
                     /transl_table=11
                     /product="ABC transport system ATP-binding protein"
                     /protein_id="CAB52348.1"
                     /db_xref="GI:5708235"
                     /translation="MDVHDVHHSYRQHAVLRGVGLRLKPGTLAGIVGENGAGKTTLLK
                     ILSGELRPDRGTVHHSGRFGYCPQTVVLDDSFTVRQHLDFFRSAFGLTGLERAEEVME
                     TLAFTEYLDQRAGRLSGGSRQKLNLTLALMHDPDVLLLDEPYQGFDWETYLRFWELAT
                     RLRDAGRSVLVVSHLAYDIDRLDQLWRLEGGRLHIQERPMEAAA"
     RBS             complement(350. .354)
                     /gene="SCJ21.02c"
     misc_feature    complement(381. .878)
                     /gene="SCJ21.02c"
                     /note="Pfam match to entry PF00005 ABC_tran, ABC
                     transporter, score 147.80, E-value 1.9e-40"
     misc_feature    complement(834. .857)
                     /gene="SCJ21.02c"
                     /note="PS00017 ATP/GTP-binding site motif A (P-loop)"
     RBS             complement(961. .964)
     RBS             1180. .1183
     misc_feature    1194. .2567
                     /gene="SCJ21.03"
                     /note="Pfam match to entry PF00743 FMO-like,
                     Flavin-binding monooxygenase-like, score 47.90, E-value
                     7.5e-25"
     gene            1194. .2570
                     /gene="SCJ21.03"
     CDS             1194. .2570
                     /gene="SCJ21.03"
                     /note="SCJ21.03, probable flavin-containing monooxygenase,
                     len: 458 aa; similar to many eukaryotic monooxygenases,
                     e.g. FMO5_RABIT dimethylaniline monooxygenase [N-oxide
                     forming] (532 aa), fasta scores; opt: 938 z-score: 1016.9
                     E(): 0, 38.7% identity in 432 aa overlap. Also similar to
                     S. coelicolor putative oxidoreductase SCH35.27 E():
                     3.4e-07, 34.8% identity in 282 aa overlap. Contains Pfam
                     match to entry PF00743 FMO-like, Flavin-binding
                     monooxygenase-like"
                     /codon_start=1
                     /transl_table=11
                     /product="putative flavin-containing monooxygenase"
                     /protein_id="CAB52349.1"
                     /db_xref="GI:5708236"
                     /translation="MRRTCVIGAGPSGLAASRVLASRGIPFDCYEAGSGIGGLWRYGN
                     DNGMSGVYASLHANISKESMSFSSLPMPDSYPVFPHHTQVLAYLESYAETFGLHGHIG
                     LRTEVTSVRPVEGGGWEVTRRSRGGAEAETDRYTEVVVANGHHWDPRLPDPAVPGAGV
                     FEGSAVHAHAYRSPEPYAGQRVLVVGMGNSGCEIAAEISRSAARTFLSARAVAHVFPK
                     MLLGRPADHWAVSRAAVLPRFLKDPGMALLLRLARGAPALYGLPEPVRRPLAAHPSTS
                     DELLVQLARGAVTAKPGIRSFGRDSVSFTDGSRETVDAVVYATGYSLSFPFLAPAVFA
                     APDGRTELYLRTVPPRLPGLFFMGLAQPAGAAFPLLEPQAEWIADLIEGEVLLPTPAE
                     MTRSIARARERHDKVYAPSYRHGIEIDICAYRRALRRELRAGRRRARRADARRPDAPR
                     PAPVSDTV"
     gene            complement(2660. .4006)
                     /gene="SCJ21.04c"
     CDS             complement(2660. .4006)
                     /gene="SCJ21.04c"
                     /note="SCJ21.04c, unknown, len: 448 aa; some similarity to
                     M. tuberculosis hypothetical protein TR:P71694
                     (EMBL:Z95844) Rv1425 (459 aa), fasta scores; opt: 150
                     z-score: 166.6 E(): 0.063, 23.9% identity in 464 aa
                     overlap"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein SCJ2.04c"
                     /protein_id="CAB52350.1"
                     /db_xref="GI:5708237"
                     /translation="MSAPPTAPADPSTGSDRPKRSVPPPRHPTHAVDRAFLDLERRRP
                     EVRWDAGGVAYLSGPPPSLADLRAYVAFRLGELPLLTSRVTGGRRPGWQPSEDFPVDR
                     HVYEVVADGPADWDTALHSALNAPFAPGAYWGVWLIHGHEPDAYALCYRFLHACQDGS
                     AAAMTFRAMLGDGEPSARPLPGGSGVFGVPRRFGAAVALATRFLARTFTVRRHRPDPA
                     FAPAGERRLWRGRVPVDTLRLIGAGRGGSAHDAHLAALAGALSVWAARSGVPLPRVTA
```

```
                VLPVDARRADEEQTWGNRCFALPVDLPVGPAPGRPDGSPDEEAPLRRLDQIMAVTRKL
                RGDTWRQAVQDLVRYMPDRPTAWYLRRMLSPRVTNVMATSMPLAEKGSLGETRVTGTA
                LLPLLVPGHLCGVGLSFFGDWAEVSFVADRALPLGELLPELWEQAVRELAESVQRG"
     gene       4632. .5993
                /gene="SCJ21.05"
     CDS        4632. .5993
                /gene="SCJ21.05"
                /note="SCJ21.05, unknown, len: 453 aa; similar in part to
                several polyketide synthases, e.g. HETM_ANASP polyketide
                synthase HETM (506 aa), fasta scores; opt: 252 z-score:
                267.0 E(): 1.6e-07, 33.7% identity in 199 aa overlap, and
                TR:Q50858 (EMBL:U24657) saframycin MX1 synthetase A (2605
                aa), fasta scores; opt: 244 z-score: 248.0 E(): 1.9e-06,
                28.5% identity in 372 aa overlap. Contains TTA Leu codon;
                possible target for regulation by bldA"
                /codon_start=1
                /transl_table=11
                /product="hypothetical SCJ21.05"
                /protein_id="CAB52351.1"
                /db_xref="GI:5708238"
                /translation="MRGAIADDDPRGTVDAVPLVSASGPPHRRCSASARPARRGTAPL
                TALGATLVHVALTGATGFLGLRLLRRLLDTHAWVTVLAHAGSGDALHRITRSLELMGA
                PSAFLARLPERLRVVETDLALPRLGLSARAFRELADSLGAIWHSAGSIHLEGDLPELR
                RTNVEGTRHVLELAAAGRAEPVVRHVSTAFVAGARRTGVAYEDELDDGAGFENAYEQS
                KYEAELLVHAWSREHGRPALVLRPSILVTDVPSHPELPSHPLQVVERILRDARRATVA
                APPTATSSPNGAPDGSSRGAHGLPGPRARPRVRTVGHPHGRLNLLPVEHAADVMVRLS
                ELPSSHGVDTYHVVHDRDVPVPTVVALLERLVPLSIDLVDAKPEDPSALETMIDFYPG
                ITAYLTHRRRFDDTRVRTLLGPAAVCAPVGLDGLWSGLAPRSGALPRPPDSPPAPAPT
                PCA"
     misc_feature 4833. .4835
                /gene="SCJ21.05"
                /note="TTA Leu codon; possible target for regulation by
                bldA"
     RBS        6052. .6055

Alignment Scores:
Pred. No.:              1.35          Length:          31717
Score:                  141.50        Matches:         63
Percent Similarity:     39.13%        Conservative:    36
Best Local Similarity:  24.90%        Mismatches:      91
Query Match:            3.80%         Indels:          63
DB:                     1             Gaps:            12

US-09-809-638-2 (1-699) x SCJ21 (1-31717)

Qy     447 LeuGluArgSerAlaHisLeuLeuAsnGluThrGlyAlaAspPheIleThrIleLeuGlu 466
           |||   |||   |||:::::::::   ::::::||||||||   :::   :::   |||
Db   23289 CTGCGGCGCGTGGCGAACGTCATCCGGAAGAGCGGCGCAGATGTGGTCGGCCTCCAGGAG 23230

Qy     467 SerAspAlaSerLysProTyrMetGlyAsnAsnAspLeuThrMetTrp------------ 482
             |||       |||   |||         :::|||         |||
Db   23229 GTCGAC------AAGCACTACTCGGCGCGCAGCGAC---------TGGGCGGACCAGCCG 23185

Qy     483 ------LeuGlyGluLysLeuGlyPheTyrThrAspPheGly----------------- 494
                |||   |||   ||||||::::::      ||||||
Db   23184 GCGGAGTTGGCGGAGTTGCTGGGTTACCACGTCGTCTTCGGCGCCAACATCGACAACAGC 23125

Qy     495 ---------ProSerThrArgTyrHisThrTrpGlyIleMetAlaLeuSerArgTyrPro 511
                   |||                :::|||         |||||||||||||||
Db   23124 CCGCCCGCGCCCGGCGGCCACCGCGTGCAGTACGGCACCGCGATCCTTTCGCGCTACCCC 23065

Qy     512 IleValLysSerGluHis---HisLeuLeuProSerProGluGlyGluIleAlaProAla 530
           |||       |||::::::      |||       ||||||      |||
Db   23064 ATCACCGCCTCGGACAACACCTGGCTGTACAAGTCGCCGGGCCAGGAACAGCGCGGTCTG 23005

Qy     531 IleThrLeuThrValAsnIleSerGlyLysLeuValAspPheValValThrHisPhe--- 549
           :::       |||::::::::::   ||||||   |||:::|||       ||||||
Db   23004 CTGCATGCCACGCTCGACGTCCACGGCAAGAAGGTGGAGTTCTACAACACGCATCTGGCC 22945

Qy     550 ---GlyAsnHisGluAspAspLeuAspArgLysLeuGlnAlaIleAlaValSerLysLeu 568
              |||:::      |||   |||   :::      |||   :::   :::
Db   22944 GCCGGCTCACAGGCCGACCGCCTGCAGCAGACGGCTCAGGTGGTGGATCTGATCGGCACG 22885

Qy     569 LeuLysSerSerSerAsnGlnValIlePheLeuGlyTyrIleThrSerAlaProGlySer 588
             |||                  |||   :::|||         :::   |||   :::
Db   22884 AGGAAGCCGGGG-----------ATCCTTGTGGGCGATTTCAACGCGCTGCCGGCCGCT 22837

Qy     589 ArgAspTyrLeuGlnLeuThrGluHisGlyAsnValLysAspIleAspSerThrAspHis 608
             :::                                 :::   :::   :::
Db   22836 CCCGAG-------------------------TCGCGGCCGCTGCAGAATGCGTACACC 22804

Qy     609 AspArgTrpCysGluTyrIleMetTyrArgGly--------------------------- 619
           |||   |||   :::   :::      ||||||
Db   22803 GATGCGTGGGCCAAGTCCCTGCACGCGCGTGGGGACGGGGCAACCTACCCCGCCCAGTCG 22744

Qy     620 ---------LeuIleArgLeuGlyTyrAlaArgIleSerHisAlaGluLeuSerAspSer 636
                     |||   |||   ||||||                  ::::::
Db   22743 CCGACTGAGCGCATCGATCTCATCTACGCGACC---------CGCAGGGTGACGCCCCTG 22693

Qy     637 GluIleGlnMetAlaLysPheArgIleProAspAspProThrAsnTyrArgAspAsnGln 656
                 |||:::   |||            |||||||||
Db   22692 GTCGCTCAGGTCCTCAAG-----------GACGACCCG--------------------- 22666

Qy     657 LysValValIleAspHisArgGluValSerGluLysIle 669
                       |||||||||   :::      |||:::
Db   22665 ---GCGGCCTCGGACCACCGTCCGCTCCTCGGCAAGGTG 22630


Search completed: October 21, 2002, 23:28:36
Job time : 3996 secs
```

ALIGNMENTS

RESULT 1
AK025164
LOCUS AK025164 2486 bp mRNA linear PRI 29-SEP-2000
DEFINITION Homo sapiens cDNA: FLJ21511 fis, clone COL05748.
ACCESSION AK025164
VERSION AK025164.1 GI:10437625
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Homo sapiens colon cDNA to mRNA, clone_lib:COL clone:COL05748.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (sites)
AUTHORS Kawabata,A., Hikiji,T., Kobatake,N., Inagaki,H., Ikema,Y., Okamoto,S., Okitani,R., Ota,T., Suzuki,Y., Obayashi,M., Nishi,T., Shibahara,T., Tanaka,T., Nakamura,Y., Isogai,T. and Sugano,S.
TITLE NEDO human cDNA sequencing project
JOURNAL Unpublished (2000)
REFERENCE 2 (bases 1 to 2486)
AUTHORS Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T., Shibahara,T., Tanaka,T. and Nakamura,Y.
TITLE Direct Submission
JOURNAL Submitted (29-AUG-2000) Sumio Sugano, Institute of Medical Science, University of Tokyo, Laboratory of Genome Structure Analysis, Human

THIS PAGE BLANK (USPTO)

```
            Genome Center; Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639,
            Japan (E-mail:cdna1@ims.u-tokyo.ac.jp, Tel:81-3-5449-5286,
            Fax:81-3-5449-5416)
COMMENT     NEDO human cDNA sequencing project supported by Ministry of
            International Trade and Industry of Japan; cDNA full insert
            sequencing: Research Association for Biotechnology; cDNA library
            construction, 5'- & 3'-end one pass sequencing: Departent of
            Virology and Human Genome Center, Institute of Medical Science,
            University of Tokyo (partly supported by Science and Technology
            Agency).
FEATURES              Location/Qualifiers
     source           1. .2486
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone="COL05748"
                      /tissue_type="colon"
                      /clone_lib="COL"
                      /note="cloning vector pME18SFL3"
     CDS              183. .2282
                      /note="unnamed protein product"
                      /codon_start=1
                      /protein_id="BAB15080.1"
                      /db_xref="GI:10437626"
                      /translation="MTSLWREILLESLLGCVSWSLYHDLGPMIYYFPLQTLELTGLEG
                      FSIAFLSPIFLTITPFWKLVNKKWMLTLLRIITIGSIASFQAPNAKLRLMVLALGVSS
                      SLIVQAVTWWSGSHLQRYLRIWGFILGQIVLVVLRIWYTSLNPIWSYQMSNKVILTLS
                      AIATLDRIGTDGDCSKPEEKKTGEVATGMASRPNWLLAGAAFGSLVFLTHWVFGEVSL
                      VSRWAVSGHPHPGPDPNPFGGAVLLCLASGLMLPSCLWFRGTGLIWWVTGTASAAGLL
                      YLHTWAAAVSGCVFAIFTASMWPQTLGHLINSGTNPGKTMTIAMIFYLLEIFFCAWCT
                      AFKFVPGGVYARERSDVLLGTMMLIIGLNMLFGPKKNLDLLLQTKNSSKVLFRKSEKY
                      MKLFLWLLVGVGLLGLGLRHKAYERKLGKVAPTKEVSAAIWPFRFGYDNEGWSSLERS
                      AHLLNETGADFITILESDASKPYMGNNDLTMWLGEKLGFYTDFGPSTRYHTWGIMALS
                      RYPIVKSEHHLLPSPEGEIAPAITLTVNISGKLVDFVVTHFGNHEDDLDRKLQAIAVS
                      KLLKSSSNQVIFLGYITSAPGSRDYLQLTEHGNVKDIDSTDHDRWCEYIMYRGLIRLG
                      YARISHAELSDSEIQMAKFRIPDDPTNYRDNQKVVIDHREVSEKIHFNPRFGSYKEGH
                      NYENNHNFHMNTPKYFL"
BASE COUNT      692 a    516 c    606 g    672 t
ORIGIN

  Query Match           99.8%;  Score 2098.2;  DB 9;  Length 2486;
  Best Local Similarity 99.9%;  Pred. No. 0;
  Matches 2100;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 ATGACCTCGCTGTGGAGAGAAATCCTCTTGGAGTCGCTGCTGGGATGTGTTTCTTGGTCT 60
Db     183 ATGACCTCGCTGTGGAGAGAAATCCTCTTGGAGTCGCTGCTGGGATGTGTTTCTTGGTCT 242

Qy      61 CTCTACCATGACCTGGGACCGATGATCTATTACTTTCCTTTGCAAACACTAGAACTCACT 120
Db     243 CTCTACCATGACCTGGGACCGATGATCTATTACTTTCCTTTGCAAACACTAGAACTCACT 302

Qy     121 GGGCTTGAAGGTTTTAGTATAGCATTTCTTTCTCCAATATTCCTAACAATTACTCCTTTC 180
Db     303 GGGCTTGAAGGTTTTAGTATAGCATTTCTTTCTCCAATATTCCTAACAATTACTCCTTTC 362

Qy     181 TGGAAATTGGTTAACAAGAAGTGGATGCTAACCCTGCTGAGGATAATCACTATTGGCAGC 240
Db     363 TGGAAATTGGTTAACAAGAAGTGGATGCTAACCCTGCTGAGGATAATCACTATTGGCAGC 422

Qy     241 ATAGCCTCCTTCCAGGCTCCAAATGCCAAACTTCGACTGATGGTTCTTGCGCTTGGGGTG 300
Db     423 ATAGCCTCCTTCCAGGCTCCAAATGCCAAACTTCGACTGATGGTTCTTGCGCTTGGGGTG 482

Qy     301 TCTTCCTCACTGATAGTGCAAGCTGTGACTTGGTGGTCGGGAAGTCATTTGCAAAGGTAC 360
Db     483 TCTTCCTCACTGATAGTGCAAGCTGTGACTTGGTGGTCAGGAAGTCATTTGCAAAGGTAC 542

Qy     361 CTCAGAATTTGGGGATTCATTTTAGGACAGATTGTTCTTGTTGTTCTACGCATATGGTAT 420
Db     543 CTCAGAATTTGGGGATTCATTTTAGGACAGATTGTTCTTGTTGTTCTACGCATATGGTAT 602

Qy     421 ACTTCACTAAACCCAATCTGGAGTTATCAGATGTCCAACAAAGTGATACTGACATTAAGT 480
Db     603 ACTTCACTAAACCCAATCTGGAGTTATCAGATGTCCAACAAAGTGATACTGACATTAAGT 662

Qy     481 GCCATAGCCACACTTGATCGTATTGGCACAGATGGTGACTGCAGTAAACCTGAAGAAAAG 540
Db     663 GCCATAGCCACACTTGATCGTATTGGCACAGATGGTGACTGCAGTAAACCTGAAGAAAAG 722

Qy     541 AAGACTGGTGAGGTAGCCACGGGGATGGCCTCTAGACCCAACTGGCTGCTGGCAGGGGCT 600
Db     723 AAGACTGGTGAGGTAGCCACGGGGATGGCCTCTAGACCCAACTGGCTGCTGGCAGGGGCT 782

Qy     601 GCTTTTGGTAGCCTTGTGTTCCTCACCCACTGGGTTTTTGGAGAAGTCTCTCTTGTTTCC 660
Db     783 GCTTTTGGTAGCCTTGTGTTCCTCACCCACTGGGTTTTTGGAGAAGTCTCTCTTGTTTCC 842

Qy     661 AGATGGGCAGTGAGTGGGCATCCACATCCAGGGCCAGATCCTAACCCATTTGGAGGTGCA 720
Db     843 AGATGGGCAGTGAGTGGGCATCCACATCCAGGGCCAGATCCTAACCCATTTGGAGGTGCA 902

Qy     721 GTACTGCTGTGCTTGGCAAGTGGATTGATGCTTCCATCTTGTTTGTGGTTTCGTGGTACT 780
Db     903 GTACTGCTGTGCTTGGCAAGTGGATTGATGCTTCCATCTTGTTTGTGGTTTCGTGGTACT 962

Qy     781 GGTTTGATCTGGTGGGTTACAGGAACAGCTTCAGCTGCGGGGCTCCTTTACCTGCACACA 840
Db     963 GGTTTGATCTGGTGGGTTACAGGAACAGCTTCAGCTGCGGGGCTCCTTTACCTGCACACA 1022

Qy     841 TGGGCAGCTGCTGTGTCTGGCTGTGTCTTCGCCATCTTTACTGCATCCATGTGGCCCCAA 900
Db    1023 TGGGCAGCTGCTGTGTCTGGCTGTGTCTTCGCCATCTTTACTGCATCCATGTGGCCCCAA 1082

Qy     901 ACACTTGGACACCTTATTAACTCAGGGACAAACCCTGGGAAAACCATGACCATTGCCATG 960
Db    1083 ACACTTGGACACCTTATTAACTCAGGGACAAACCCTGGGAAAACCATGACCATTGCCATG 1142

Qy     961 ATATTTTATCTTCTAGAAATATTTTTCTGTGCCTGGTGCACAGCTTTTAAGTTTGTCCCA 1020
Db    1143 ATATTTTATCTTCTAGAAATATTTTTCTGTGCCTGGTGCACAGCTTTTAAGTTTGTCCCA 1202

Qy    1021 GGAGGTGTCTACGCTAGAGAAAGATCAGATGTGCTTTTGGGGACAATGATGTTAATTATC 1080
Db    1203 GGAGGTGTCTACGCTAGAGAAAGATCAGATGTGCTTTTGGGGACAATGATGTTAATTATC 1262

Qy    1081 GGGCTGAATATGCTATTTGGTCCTAAGAAAAACCTTGACTTGCTTCTTCAAACAAAAAAC 1140
Db    1263 GGGCTGAATATGCTATTTGGTCCTAAGAAAAACCTTGATTTGCTTCTTCAAACAAAAAAC 1322

Qy    1141 AGTTCTAAAGTGCTTTTCAGAAAGAGTGAAAAATACATGAAACTTTTTCTGTGGCTGCTT 1200
Db    1323 AGTTCTAAAGTGCTTTTCAGAAAGAGTGAAAAATACATGAAACTTTTTCTGTGGCTGCTT 1382

Qy    1201 GTTGGTGTGGGATTGTTGGGATTAGGACTACGGCATAAAGCCTATGAGAGAAAACTGGGC 1260
Db    1383 GTTGGTGTGGGATTGTTGGGATTAGGACTACGGCATAAAGCCTATGAGAGAAAACTGGGC 1442

Qy    1261 AAAGTGGCACCAACCAAAGAGGTCTCTGCTGCCATCTGGCCTTTCAGGTTTGGATATGAC 1320
Db    1443 AAAGTGGCACCAACCAAAGAGGTCTCTGCTGCCATCTGGCCTTTCAGGTTTGGATATGAC 1502

Qy    1321 AATGAAGGGTGGTCTAGTCTAGAAAGATCAGCTCACCTGCTCAATGAAACAGGTGCAGAT 1380
Db    1503 AATGAAGGGTGGTCTAGTCTAGAAAGATCAGCTCACCTGCTCAATGAAACAGGTGCAGAT 1562

Qy    1381 TTCATAACAATTTTGGAGAGTGATGCTTCTAAGCCCTATATGGGGAACAATGACTTAACC 1440
Db    1563 TTCATAACAATTTTGGAGAGTGATGCTTCTAAGCCCTATATGGGGAACAATGACTTAACC 1622

Qy    1441 ATGTGGCTAGGGGAAAAGTTGGGTTTCTATACAGACTTTGGTCCAAGCACAAGGTATCAC 1500
Db    1623 ATGTGGCTAGGGGAAAAGTTGGGTTTCTATACAGACTTTGGTCCAAGCACAAGGTATCAC 1682

Qy    1501 ACTTGGGGGATTATGGCTTTGTCAAGATACCCAATTGTGAAATCTGAGCATCACCTTCTT 1560
Db    1683 ACTTGGGGGATTATGGCTTTGTCAAGATACCCAATTGTGAAATCTGAGCATCACCTTCTT 1742
```

[redacted] Oct 23 09:09:

THIS PAGE BLANK (USPTO)

```
Qy   1561 CCGTCACCAGAGGGCGAGATCGCACCAGCCATCACATTGACCGTTAACATTTCGGGCAAG 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1743 CCGTCACCAGAGGGCGAGATCGCACCAGCCATCACATTGACCGTTAACATTTCGGGCAAG 1802

Qy   1621 CTGGTGGATTTTGTCGTGACACACTTTGGGAACCACGAAGATGACCTCGACAGGAAACTG 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1803 CTGGTGGATTTTGTCGTGACACACTTTGGGAACCACGAAGATGACCTCGACAGGAAACTG 1862

Qy   1681 CAGGCTATTGCTGTTTCAAAACTACTGAAAAGTAGCTCTAATCAAGTGATATTTCTGGGA 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1863 CAGGCTATTGCTGTTTCAAAACTACTGAAAAGTAGCTCTAATCAAGTGATATTTCTGGGA 1922

Qy   1741 TATATCACTTCAGCACCTGGCTCCAGAGATTATCTACAGCTCACTGAACATGGCAATGTG 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1923 TATATCACTTCAGCACCTGGCTCCAGAGATTATCTACAGCTCACTGAACATGGCAATGTG 1982

Qy   1801 AAGGATATCGACAGCACTGATCATGACAGATGGTGTGAATACATTATGTATCGAGGGCTG 1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1983 AAGGATATCGACAGCACTGATCATGACAGATGGTGTGAATACATTATGTATCGAGGGCTG 2042

Qy   1861 ATCAGGTTGGGTTATGCAAGAATCTCCCATGCTGAACTGAGTGATTCAGAAATTCAGATG 1920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2043 ATCAGGTTGGGTTATGCAAGAATCTCCCATGCTGAACTGAGTGATTCAGAAATTCAGATG 2102

Qy   1921 GCAAAATTTAGGATCCCTGATGACCCCACTAATTATAGAGACAACCAGAAAGTGGTCATA 1980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2103 GCAAAATTTAGGATCCCTGATGACCCCACTAATTATAGAGACAACCAGAAAGTGGTCATA 2162

Qy   1981 GACCACAGAGAAGTTTCTGAGAAAATTCATTTTAATCCCAGATTTGGATCCTACAAAGAA 2040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2163 GACCACAGAGAAGTTTCTGAGAAAATTCATTTTAATCCCAGATTTGGATCCTACAAAGAA 2222

Qy   2041 GGACACAATTATGAAAACAACCATCATTTTCATATGAATACTCCCAAATACTTTTTATGA 2100
          |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db   2223 GGACACAATTATGAAAACAACCATAATTTTCATATGAATACTCCCAAATACTTTTTATGA 2282

Qy   2101 AAC 2103
          |||
Db   2283 AAC 2285
```

RESULT 2
BC016523
LOCUS
D[redacted]

GenCore version 5.1.3

OM nucleic - nucleic search, using sw model

Run on: October 21, 2002, 18:04:48 ; Search time 3823 Seconds
(without alignments)
11511.511 Million cell updates/sec

Title: US-09-809-638-1
Perfect score: 2103
Sequence: 1 atgacctcgctgtggagaga.....ccaaatacttttttatgaaac 2103

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htgo_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	2098.2	99.8	2486	9	AK025164	AK025164 Homo sapi
2	1511	71.8	2412	10	BC016523	BC016523 Mus muscu
3	267.4	12.7	184206	9	AC020593	AC020593 Homo sapi
4	160	7.6	222885	2	AL158816	AL158816 Homo sapi
c 5	151	7.2	222885	2	AL158816	AL158816 Homo sapi
c 6	132.4	6.3	316613	8	SCCHRIII	X59720 S.cerevisia
7	121.8	5.8	1234	8	D89130	D89130 Schizosacch
8	121.8	5.8	34034	8	SPAC688	AL355632 S.pombe c
c 9	112	5.3	305	6	AX337478	AX337478 Sequence
10	110.6	5.3	995	11	CNS06JWV	AL402149 T3 end of
c 11	81.6	3.9	165893	2	AC106576	AC106576 Rattus no
c 12	81.4	3.9	1348	9	HUMTPIPSA	K03223 Human TPI (
c 13	54	2.6	54	6	AX106346	AX106346 Sequence
c 14	54	2.6	54	6	AX140637	AX140637 Sequence
c 15	54	2.6	54	6	AX200497	AX200497 Sequence
c 16	54	2.6	54	6	AX267153	AX267153 Sequence
c 17	42.6	2.0	96374	2	AC095821	AC095821 Rattus no
18	42.4	2.0	151931	30	AC011956	Ac011956 Homo sapi
c 19	42	2.0	162349	9	AC006226	AC006226 Homo sapi
20	42	2.0	190000	2	AC007072	AC007072 Homo sapi
c 21	42	2.0	199528	2	AC108864	AC108864 Homo sapi
22	41.4	2.0	167750	2	AC016439	AC016439 Homo sapi
23	41.4	2.0	172710	2	AC021928	AC021928 Homo sapi
c 24	41.2	2.0	141577	9	AL606748	AL606748 Human DNA
25	41	1.9	6563	10	RATCYP450A	M26129 Rat cytochr
c 26	41	1.9	209712	2	AL627072	AL627072 Mus muscu
c 27	40.8	1.9	408	11	G51484	G51484 SHGC-82855
c 28	40.8	1.9	42114	3	U50193	U50193 Caenorhabdi
c 29	40.8	1.9	124067	9	AP000435	AP000435 Homo sapi
c 30	40.8	1.9	147358	9	AP002345	AP002345 Homo sapi
31	40.6	1.9	94947	2	AC099240	AC099240 Rattus no
32	40.4	1.9	100061	2	AC016551	AC016551 Homo sapi
33	40.4	1.9	170569	2	AC108109	AC108109 Homo sapi
c 34	40.4	1.9	191132	9	AC010282	AC010282 Homo sapi
c 35	40.2	1.9	34568	2	AC083896	AC083896 Homo sapi
36	40.2	1.9	96280	9	AC006968	AC006968 Homo sapi
c 37	40.2	1.9	209524	2	AC099602	AC099602 Mus muscu
38	39.8	1.9	1141	6	AX083744	AX083744 Sequence
39	39.8	1.9	54814	2	AC104718	AC104718 Sus scrof
c 40	39.8	1.9	67460	2	AC102147	AC102147 Ciona sav
41	39.6	1.9	6917	10	RATCYP45C	K02246 Rat cytochr
42	39.4	1.9	20250	10	AF260315	AF260315 Mus muscu
43	39.4	1.9	163782	9	AC011228	AC011228 Homo sapi
44	39.4	1.9	177402	9	AC019210	AC019210 Homo sapi
c 45	39.2	1.9	71229	2	AL451104	AL451104 Homo sapi

ALIGNMENTS

RESULT 1
AK025164
LOCUS AK025164 2486 bp mRNA linear PRI 29-SEP-2000
DEFINITION Homo sapiens cDNA: FLJ21511 fis, clone COL05748.
ACCESSION AK025164
VERSION AK025164.1 GI:10437625
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Homo sapiens colon cDNA to mRNA, clone_lib:COL clone:COL05748.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (sites)
AUTHORS Kawabata,A., Hikiji,T., Kobatake,N., Inagaki,H., Ikema,Y., Okamoto,S., Okitani,R., Ota,T., Suzuki,Y., Obayashi,M., Nishi,T., Shibahara,T., Tanaka,T., Nakamura,Y., Isogai,T. and Sugano,S.
TITLE NEDO human cDNA sequencing project
JOURNAL Unpublished (2000)
REFERENCE 2 (bases 1 to 2486)
AUTHORS Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T., Shibahara,T., Tanaka,T. and Nakamura,Y.
TITLE Direct Submission
JOURNAL Submitted (29-AUG-2000) Sumio Sugano, Institute of Medical Science, University of Tokyo, Laboratory of Genome Structure Analysis, Human

Genome Center; Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan (E-mail:cdna1@ims.u-tokyo.ac.jp, Tel:81-3-5449-5286, Fax:81-3-5449-5416)

COMMENT NEDO human cDNA sequencing project supported by Ministry of International Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology; cDNA library construction, 5'- & 3'-end one pass sequencing: Departent of Virology and Human Genome Center, Institute of Medical Science, University of Tokyo (partly supported by Science and Technology Agency).

FEATURES Location/Qualifiers

source 1. .2486
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="COL05748"
/tissue_type="colon"
/clone_lib="COL"
/note="cloning vector pME18SFL3"

CDS 183. .2282
/note="unnamed protein product"
/codon_start=1
/protein_id="BAB15080.1"
/db_xref="GI:10437626"
/translation="MTSLWREILLESLLGCVSWSLYHDLGPMIYYFPLQTLELTGLEG
FSIAFLSPIFLTITPFWKLVNKKWMLTLLRIITIGSIASFQAPNAKLRLMVLALGVSS
SLIVQAVTWWSGSHLQRYLRIWGFILGQIVLVVLRIWYTSLNPIWSYQMSNKVILTLS
AIATLDRIGTDGDCSKPEEKKTGEVATGMASRPNWLLAGAAFGSLVFLTHWVFGEVSL
VSRWAVSGHPHPGPDPNPFGGAVLLCLASGLMLPSCLWFRGTGLIWWVTGTASAAGLL
YLHTWAAAVSGCVFAIFTASMWPQTLGHLINSGTNPGKTMTIAMIFYLLEIFFCAWCT
AFKFVPGGVYARERSDVLLGTMMLIIGLNMLFGPKKNLDLLLQTKNSSKVLFRKSEKY
MKLFLWLLVGVGLLGLGLRHKAYERKLGKVAPTKEVSAAIWPFRFGYDNEGWSSLERS
AHLLNETGADFITILESDASKPYMGNNDLTMWLGEKLGFYTDFGPSTRYHTWGIMALS
RYPIVKSEHHLLPSPEGEIAPAITLTVNISGKLVDFVVTHFGNHEDDLDRKLQAIAVS
KLLKSSSNQVIFLGYITSAPGSRDYLQLTEHGNVKDIDSTDHDRWCEYIMYRGLIRLG
YARISHAELSDSEIQMAKFRIPDDPTNYRDNQKVVIDHREVSEKIHFNPRFGSYKEGH
NYENNHNFHMNTPKYFL"

BASE COUNT 692 a 516 c 606 g 672 t

ORIGIN

Query Match 99.8%; Score 2098.2; DB 9; Length 2486;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2100; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
Qy       1 ATGACCTCGCTGTGGAGAGAAATCCTCTTGGAGTCGCTGCTGGGATGTGTTTCTTGGTCT 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     183 ATGACCTCGCTGTGGAGAGAAATCCTCTTGGAGTCGCTGCTGGGATGTGTTTCTTGGTCT 242

Qy      61 CTCTACCATGACCTGGGACCGATGATCTATTACTTTCCTTTGCAAACACTAGAACTCACT 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     243 CTCTACCATGACCTGGGACCGATGATCTATTACTTTCCTTTGCAAACACTAGAACTCACT 302

Qy     121 GGGCTTGAAGGTTTTAGTATAGCATTTCTTTCTCCAATATTCCTAACAATTACTCCTTTC 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     303 GGGCTTGAAGGTTTTAGTATAGCATTTCTTTCTCCAATATTCCTAACAATTACTCCTTTC 362

Qy     181 TGGAAATTGGTTAACAAGAAGTGGATGCTAACCCTGCTGAGGATAATCACTATTGGCAGC 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     363 TGGAAATTGGTTAACAAGAAGTGGATGCTAACCCTGCTGAGGATAATCACTATTGGCAGC 422

Qy     241 ATAGCCTCCTTCCAGGCTCCAAATGCCAAACTTCGACTGATGGTTCTTGCGCTTGGGGTG 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     423 ATAGCCTCCTTCCAGGCTCCAAATGCCAAACTTCGACTGATGGTTCTTGCGCTTGGGGTG 482

Qy     301 TCTTCCTCACTGATAGTGCAAGCTGTGACTTGGTGGTCGGGAAGTCATTTGCAAAGGTAC 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     483 TCTTCCTCACTGATAGTGCAAGCTGTGACTTGGTGGTCAGGAAGTCATTTGCAAAGGTAC 542

Qy     361 CTCAGAATTTGGGGATTCATTTTAGGACAGATTGTTCTTGTTGTTCTACGCATATGGTAT 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     543 CTCAGAATTTGGGGATTCATTTTAGGACAGATTGTTCTTGTTGTTCTACGCATATGGTAT 602

Qy     421 ACTTCACTAAACCCAATCTGGAGTTATCAGATGTCCAACAAAGTGATACTGACATTAAGT 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     603 ACTTCACTAAACCCAATCTGGAGTTATCAGATGTCCAACAAAGTGATACTGACATTAAGT 662

Qy     481 GCCATAGCCACACTTGATCGTATTGGCACAGATGGTGACTGCAGTAAACCTGAAGAAAAG 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     663 GCCATAGCCACACTTGATCGTATTGGCACAGATGGTGACTGCAGTAAACCTGAAGAAAAG 722

Qy     541 AAGACTGGTGAGGTAGCCACGGGGATGGCCTCTAGACCCAACTGGCTGCTGGCAGGGGCT 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     723 AAGACTGGTGAGGTAGCCACGGGGATGGCCTCTAGACCCAACTGGCTGCTGGCAGGGGCT 782

Qy     601 GCTTTTGGTAGCCTTGTGTTCCTCACCCACTGGGTTTTTGGAGAAGTCTCTCTTGTTTCC 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     783 GCTTTTGGTAGCCTTGTGTTCCTCACCCACTGGGTTTTTGGAGAAGTCTCTCTTGTTTCC 842

Qy     661 AGATGGGCAGTGAGTGGGCATCCACATCCAGGGCCAGATCCTAACCCATTTGGAGGTGCA 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     843 AGATGGGCAGTGAGTGGGCATCCACATCCAGGGCCAGATCCTAACCCATTTGGAGGTGCA 902

Qy     721 GTACTGCTGTGCTTGGCAAGTGGATTGATGCTTCCATCTTGTTTGTGGTTTCGTGGTACT 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     903 GTACTGCTGTGCTTGGCAAGTGGATTGATGCTTCCATCTTGTTTGTGGTTTCGTGGTACT 962

Qy     781 GGTTTGATCTGGTGGGTTACAGGAACAGCTTCAGCTGCGGGGCTCCTTTACCTGCACACA 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     963 GGTTTGATCTGGTGGGTTACAGGAACAGCTTCAGCTGCGGGGCTCCTTTACCTGCACACA 1022

Qy     841 TGGGCAGCTGCTGTGTCTGGCTGTGTCTTCGCCATCTTTACTGCATCCATGTGGCCCCAA 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1023 TGGGCAGCTGCTGTGTCTGGCTGTGTCTTCGCCATCTTTACTGCATCCATGTGGCCCCAA 1082

Qy     901 ACACTTGGACACCTTATTAACTCAGGGACAAACCCTGGGAAAACCATGACCATTGCCATG 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1083 ACACTTGGACACCTTATTAACTCAGGGACAAACCCTGGGAAAACCATGACCATTGCCATG 1142

Qy     961 ATATTTTATCTTCTAGAAATATTTTTCTGTGCCTGGTGCACAGCTTTTAAGTTTGTCCCA 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1143 ATATTTTATCTTCTAGAAATATTTTTCTGTGCCTGGTGCACAGCTTTTAAGTTTGTCCCA 1202

Qy    1021 GGAGGTGTCTACGCTAGAGAAAGATCAGATGTGCTTTTGGGGACAATGATGTTAATTATC 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1203 GGAGGTGTCTACGCTAGAGAAAGATCAGATGTGCTTTTGGGGACAATGATGTTAATTATC 1262

Qy    1081 GGGCTGAATATGCTATTTGGTCCTAAGAAAAACCTTGACTTGCTTCTTCAAACAAAAAAC 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1263 GGGCTGAATATGCTATTTGGTCCTAAGAAAAACCTTGATTTGCTTCTTCAAACAAAAAAC 1322

Qy    1141 AGTTCTAAAGTGCTTTTCAGAAAGAGTGAAAAATACATGAAACTTTTTCTGTGGCTGCTT 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1323 AGTTCTAAAGTGCTTTTCAGAAAGAGTGAAAAATACATGAAACTTTTTCTGTGGCTGCTT 1382

Qy    1201 GTTGGTGTGGGATTGTTGGGATTAGGACTACGGCATAAAGCCTATGAGAGAAAACTGGGC 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1383 GTTGGTGTGGGATTGTTGGGATTAGGACTACGGCATAAAGCCTATGAGAGAAAACTGGGC 1442

Qy    1261 AAAGTGGCACCAACCAAAGAGGTCTCTGCTGCCATCTGGCCTTTCAGGTTTGGATATGAC 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1443 AAAGTGGCACCAACCAAAGAGGTCTCTGCTGCCATCTGGCCTTTCAGGTTTGGATATGAC 1502

Qy    1321 AATGAAGGGTGGTCTAGTCTAGAAAGATCAGCTCACCTGCTCAATGAAACAGGTGCAGAT 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1503 AATGAAGGGTGGTCTAGTCTAGAAAGATCAGCTCACCTGCTCAATGAAACAGGTGCAGAT 1562

Qy    1381 TTCATAACAATTTTGGAGAGTGATGCTTCTAAGCCCTATATGGGGAACAATGACTTAACC 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1563 TTCATAACAATTTTGGAGAGTGATGCTTCTAAGCCCTATATGGGGAACAATGACTTAACC 1622

Qy    1441 ATGTGGCTAGGGGAAAAGTTGGGTTTCTATACAGACTTTGGTCCAAGCACAAGGTATCAC 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1623 ATGTGGCTAGGGGAAAAGTTGGGTTTCTATACAGACTTTGGTCCAAGCACAAGGTATCAC 1682

Qy    1501 ACTTGGGGGATTATGGCTTTGTCAAGATACCCAATTGTGAAATCTGAGCATCACCTTCTT 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1683 ACTTGGGGGATTATGGCTTTGTCAAGATACCCAATTGTGAAATCTGAGCATCACCTTCTT 1742
```

```
Qy   1561 CCGTCACCAGAGGGCGAGATCGCACCAGCCATCACATTGACCGTTAACATTTCGGGCAAG 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1743 CCGTCACCAGAGGGCGAGATCGCACCAGCCATCACATTGACCGTTAACATTTCGGGCAAG 1802

Qy   1621 CTGGTGGATTTTGTCGTGACACACTTTGGGAACCACGAAGATGACCTCGACAGGAAACTG 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1803 CTGGTGGATTTTGTCGTGACACACTTTGGGAACCACGAAGATGACCTCGACAGGAAACTG 1862

Qy   1681 CAGGCTATTGCTGTTTCAAAACTACTGAAAAGTAGCTCTAATCAAGTGATATTTCTGGGA 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1863 CAGGCTATTGCTGTTTCAAAACTACTGAAAAGTAGCTCTAATCAAGTGATATTTCTGGGA 1922

Qy   1741 TATATCACTTCAGCACCTGGCTCCAGAGATTATCTACAGCTCACTGAACATGGCAATGTG 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1923 TATATCACTTCAGCACCTGGCTCCAGAGATTATCTACAGCTCACTGAACATGGCAATGTG 1982

Qy   1801 AAGGATATCGACAGCACTGATCATGACAGATGGTGTGAATACATTATGTATCGAGGGCTG 1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1983 AAGGATATCGACAGCACTGATCATGACAGATGGTGTGAATACATTATGTATCGAGGGCTG 2042

Qy   1861 ATCAGGTTGGGTTATGCAAGAATCTCCCATGCTGAACTGAGTGATTCAGAAATTCAGATG 1920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2043 ATCAGGTTGGGTTATGCAAGAATCTCCCATGCTGAACTGAGTGATTCAGAAATTCAGATG 2102

Qy   1921 GCAAAATTTAGGATCCCTGATGACCCCACTAATTATAGAGACAACCAGAAAGTGGTCATA 1980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2103 GCAAAATTTAGGATCCCTGATGACCCCACTAATTATAGAGACAACCAGAAAGTGGTCATA 2162

Qy   1981 GACCACAGAGAAGTTTCTGAGAAAATTCATTTTAATCCCAGATTTGGATCCTACAAAGAA 2040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2163 GACCACAGAGAAGTTTCTGAGAAAATTCATTTTAATCCCAGATTTGGATCCTACAAAGAA 2222

Qy   2041 GGACACAATTATGAAAACAACCATCATTTTCATATGAATACTCCCAAATACTTTTTATGA 2100
          ||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||||
Db   2223 GGACACAATTATGAAAACAACCATAATTTTCATATGAATACTCCCAAATACTTTTTATGA 2282

Qy   2101 AAC 2103
          |||
Db   2283 AAC 2285
```

RESULT 2
BC016523
LOCUS BC016523 2412 bp mRNA linear ROD 05-NOV-2001
DEFINITION Mus musculus, Similar to hypothetical protein FLJ21511, clone MGC:27925 IMAGE:3584006, mRNA, complete cds.
ACCESSION BC016523
VERSION BC016523.1 GI:16741399
KEYWORDS MGC.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 2412)
AUTHORS Strausberg,R.
TITLE Direct Submission
JOURNAL Submitted (31-OCT-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
REMARK NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT Contact: MGC help desk
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Baylor College of Medicine Human Genome Sequencing Center
Center code: BCM-HGSC
Web site: http://www.hgsc.bcm.tmc.edu/cdna/
Contact: villalon@bcm.tmc.edu.
Villalon, D.K., Luna, R.A., Hale, S.M., Hulyk, S., Lu, X., Garcia, A.M., Holloway, M., Telford, B, Hodgson, A., Bouck, J., Yu, W., Muzny,D.M., Gibbs,R.A.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAK Plate: 35 Row: k Column: 5
This clone was selected for full length sequencing because it passed the following selection criteria: Similarity but not identity to protein.

FEATURES Location/Qualifiers
source 1. .2412
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="MGC:27925 IMAGE:3584006"
/tissue_type="Mammary tumor. Metallothionien-TGF alpha model. 10 month old virgin mouse. Taken by biopsy."
/clone_lib="NCI_CGAP_Mam1"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"
CDS 113. .2212
/codon_start=1
/product="Similar to hypothetical protein FLJ21511"
/protein_id="AAH16523.1"
/db_xref="GI:16741400"
/translation="MPGLWRAIALETLLGYVSWSLYHGLSPMIYYFPLQTLELTGLEF
FCVAFLSPILLTIPPLWKLVNKKWTLSLLRIVTVGSIASFEAPNAKLRLMVLALGVSS
SLIVQTVTWWSGSGLQRYLKIWGFILGHVLLLVLRIWYTSLNPIWSYQMSNRVILTLS
AVAVLDRIGTDGDYRNPEGKKPREVATGRTSLSSWLLPGAAFGSLLFLTHWIFGEVSI
VSRWAVSGHPHPGPDPNPFGGAVLLGFSSGLMLSGSSWLHDAGLAWWMTGAASAMGLL
YLRTWAAAVSGCVLAVFTGSMWPQVLGHLVNSGKNSGEAMATGMILYVLQTFFCAWCT
AFKFVPGGVYARERSDVLLGTIMVIIGLSMLFGPKKNLDFLLQTKNSPKTLLRCSEKY
MKLILWLFVGVGLLGLGLRHRTYERQLGRGAPATVVSAAIWPFRFGYDNEGWPNLERS
AQLLKETGADFITILESDASKPYIGNNDLTMWLGEKLGFYTDFGPSTRDHTWGIMVLS
RYPIVRSEHHLLPSPEGEIAPAITMTVNVSNRLVDFVVTHFGNHEDDLDRKLQAIAVS
KLLKNCSNQVIFLGYITSEPGSRDYIQLTKHGNVKDIDSSDGDRWCEYIMYRGLIRLG
YARISHAELSDSEIQMAKFRIPDDPANYRDNQKVVIDHRGVPKNIHFNPRFGSYKEGH
NYENTHHFHMNTPKYFV"
BASE COUNT 628 a 561 c 616 g 607 t
ORIGIN

Query Match 71.8%; Score 1511; DB 10; Length 2412;
Best Local Similarity 82.4%; Pred. No. 0;
Matches 1733; Conservative 0; Mismatches 370; Indels 0; Gaps 0;

```
Qy      1 ATGACCTCGCTGTGGAGAGAAATCCTCTTGGAGTCGCTGCTGGGATGTGTTTCTTGGTCT 60
          ||| |    |||||||||| |||   | ||||| |  ||||||| |  || || |||||||
Db    113 ATGCCAGGCCTGTGGAGAGCAATTGCCCTGGAGACCTTGCTGGGCTACGTCTCCTGGTCT 172

Qy     61 CTCTACCATGACCTGGGACCGATGATCTATTACTTTCCTTTGCAAACACTAGAACTCACT 120
          |||||||| |  ||| | || |||||||||| |||||  |||||||  || || |||||
Db    173 CTCTACCACGGACTGAGCCCCATGATCTATTATTTTCCCCTGCAAACGCTGGAGCTCACA 232

Qy    121 GGGCTTGAAGGTTTTAGTATAGCATTTCTTTCTCCAATATTCCTAACAATTACTCCTTTC 180
          |||||  |||   ||| || |||| ||||| ||||| ||| || |||||||| | ||  ||
Db    233 GGGCTGGAATTCTTTTGTGTAGCCTTTCTCTCTCCGATACTCTTAACAATTCCCCCACTC 292

Qy    181 TGGAAATTGGTTAACAAGAAGTGGATGCTAACCCTGCTGAGGATAATCACTATTGGCAGC 240
          ||||||  | || || ||||||||||  ||  | ||||||||||  |||||  | ||||||
Db    293 TGGAAACTCGTCAATAAGAAGTGGACACTCTCTCTGCTGAGGATCGTCACTGTAGGCAGC 352

Qy    241 ATAGCCTCCTTCCAGGCTCCAAATGCCAAACTTCGACTGATGGTTCTTGCGCTTGGGGTG 300
          |||||  |||||  ||||||||||||||| ||||||||||||| || || |||||  |||
Db    353 ATAGCTTCCTTTGAGGCTCCAAATGCCAAGCTTCGACTGATGGTCCTCGCCCTTGGTGTG 412

Qy    301 TCTTCCTCACTGATAGTGCAAGCTGTGACTTGGTGGTCGGGAAGTCATTTGCAAAGGTAC 360
          |||||  ||  |||||||||||| ||||||| |||||||  ||||||   ||||||||||||
Db    413 TCTTCTTCCTTGATAGTGCAAACTGTGACCTGGTGGTCAGGAAGTGGATTGCAAAGGTAC 472

Qy    361 CTCAGAATTTGGGGATTCATTTTAGGACAGATTGTTCTTGTTGTTCTACGCATATGGTAT 420
          |||| ||| ||||||||||||| | |||||  |  ||||| |||| || || ||||||||
Db    473 CTCAAAATCTGGGGATTCATTCTGGGACACGTCCTTCTTCTTGTCCTCCGAATATGGTAC 532

Qy    421 ACTTCACTAAACCCAATCTGGAGTTATCAGATGTCCAACAAAGTGATACTGACATTAAGT 480
```

```
Db    533 ACATCATTAAACCCAATTTGGAGCTATCAGATGTCCAACAGAGTGATTCTGACATTGAGT 592

Qy    481 GCCATAGCCACACTTGATCGTATTGGCACAGATGGTGACTGCAGTAAACCTGAAGAAAAG 540

Db    593 GCCGTAGCTGTGCTTGACCGAATTGGCACAGATGGTGACTACAGAAACCCTGAGGGGAAG 652

Qy    541 AAGACTGGTGAGGTAGCCACGGGGATGGCCTCTAGACCCAACTGGCTGCTGGCAGGGGCT 600

Db    653 AAGCCCCGTGAGGTTGCCACGGGCAGGACCTCTCTATCTAGTTGGCTGTTGCCAGGCGCC 712

Qy    601 GCTTTTGGTAGCCTTGTGTTCCTCACCCACTGGGTTTTTGGAGAAGTCTCTCTTGTTTCC 660

Db    713 GCCTTTGGGAGCCTCCTGTTCCTAACTCACTGGATCTTCGGAGAGGTCTCCATTGTTTCC 772

Qy    661 AGATGGGCAGTGAGTGGGCATCCACATCCAGGGCCAGATCCTAACCCATTTGGAGGTGCA 720

Db    773 AGATGGGCAGTCAGTGGTCATCCCCACCCAGGGCCAGATCCTAATCCTTTTGGAGGTGCG 832

Qy    721 GTACTGCTGTGCTTGGCAAGTGGATTGATGCTTCCATCTTGTTTGTGGTTTCGTGGTACT 780

Db    833 GTTCTACTGGGCTTCTCAAGTGGACTGATGCTATCGGGTTCATCATGGCTTCATGATGCT 892

Qy    781 GGTTTGATCTGGTGGGTTACAGGAACAGCTTCAGCTGCGGGGCTCCTTTACCTGCACACA 840

Db    893 GGGTTAGCCTGGTGGATGACAGGAGCAGCTTCGGCCATGGGTCTCCTCTACCTGCGCACA 952

Qy    841 TGGGCAGCTGCTGTGTCTGGCTGTGTCTTCGCCATCTTTACTGCATCCATGTGGCCCCAA 900

Db    953 TGGGCAGCTGCTGTTTCTGGATGTGTCCTGGCAGTCTTTACGGGGTCTATGTGGCCTCAA 1012

Qy    901 ACACTTGGACACCTTATTAACTCAGGGACAAACCCTGGGAAAACCATGACCATTGCCATG 960

Db    1013 GTACTTGGCCATCTTGTGAACTCAGGGAAAAACTCCGGAGAAGCCATGGCCACTGGCATG 1072

Qy    961 ATATTTTATCTTCTAGAAATATTTTTCTGTGCCTGGTGCACAGCTTTTAAGTTTGTCCCA 1020

Db    1073 ATCCTTTATGTTTTACAAACATTTTTCTGTGCCTGGTGCACAGCGTTTAAGTTTGTTCCC 1132

Qy    1021 GGAGGTGTCTACGCTAGAGAAAGATCAGATGTGCTTTTGGGGACAATGATGTTAATTATC 1080

Db    1133 GGAGGTGTCTACGCTAGAGAAAGATCTGATGTGCTTTTGGGGACCATAATGGTAATTATT 1192

Qy    1081 GGGCTGAATATGCTATTTGGTCCTAAGAAAAACCTTGACTTGCTTCTTCAAACAAAAAAC 1140

Db    1193 GGATTGAGTATGCTGTTTGGACCTAAGAAAAACCTTGACTTTCTTCTTCAAACAAAAAAC 1252

Qy    1141 AGTTCTAAAGTGCTTTTCAGAAAGAGTGAAAAATACATGAAACTTTTTCTGTGGCTGCTT 1200

Db    1253 AGTCCTAAAACGCTTTTGAGATGCAGTGAAAAATACATGAAACTCATTCTGTGGCTGTTT 1312

Qy    1201 GTTGGTGTGGGATTGTTGGGATTAGGACTACGGCATAAAGCCTATGAGAGAAAACTGGGC 1260

Db    1313 GTTGGTGTGGGGTTACTGGGATTAGGGCTACGGCACAGAACCTACGAGAGGCAGTTAGGC 1372

Qy    1261 AAAGTGGCACCAACCAAAGAGGTCTCTGCTGCCATCTGGCCTTTCAGGTTTGGATATGAC 1320

Db    1373 CGAGGGGCACCAGCGACAGTGGTCTCGGCTGCCATCTGGCCTTTTAGGTTTGGATATGAC 1432

Qy    1321 AATGAAGGGTGGTCTAGTCTAGAAAGATCAGCTCACCTGCTCAATGAAACAGGTGCAGAT 1380

Db    1433 AATGAAGGCTGGCCCAATCTAGAGAGGTCTGCTCAACTGCTCAAGGAGACAGGTGCAGAT 1492

Qy    1381 TTCATAACAATTTTGGAGAGTGATGCTTCTAAGCCCTATATGGGGAACAATGACTTAACC 1440

Db    1493 TTTATCACCATTTTGGAGAGCGATGCTTCCAAGCCCTATATAGGGAACAACGACTTAACA 1552

Qy    1441 ATGTGGCTAGGGGAAAAGTTGGGTTTCTATACAGACTTTGGTCCAAGCACAAGGTATCAC 1500

Db    1553 ATGTGGCTGGGGGAGAAGCTGGGTTTCTACACAGACTTTGGGCCAAGCACCAGGGATCAC 1612

Qy    1501 ACTTGGGGGATTATGGCTTTGTCAAGATACCCAATTGTGAAATCTGAGCATCACCTTCTT 1560

Db    1613 ACCTGGGGGATTATGGTGCTGTCTCGGTACCCGATTGTCAGATCGGAACATCACCTTCTT 1672

Qy    1561 CCGTCACCAGAGGGCGAGATCGCACCAGCCATCACATTGACCGTTAACATTTCGGGCAAG 1620

Db    1673 CCGTCGCCAGAGGGCGAGATCGCACCAGCCATAACCATGACAGTTAACGTCTCCAACAGA 1732

Qy    1621 CTGGTGGATTTTGTCGTGACACACTTTGGGAACCACGAAGATGACCTCGACAGGAAACTG 1680

Db    1733 CTGGTGGATTTTGTGGTGACACACTTTGGGAATCATGAAGATGACCTTGACAGGAAGCTA 1792

Qy    1681 CAGGCTATTGCTGTTTCAAAACTACTGAAAAGTAGCTCTAATCAAGTGATATTTCTGGGA 1740

Db    1793 CAGGCTATTGCAGTTTCAAAACTGCTGAAAAATTGTTCAAATCAAGTGATATTTCTGGGT 1852

Qy    1741 TATATCACTTCAGCACCTGGCTCCAGAGATTATCTACAGCTCACTGAACATGGCAATGTG 1800

Db    1853 TATATCACTTCAGAGCCTGGTTCCAGAGATTACATACAACTCACAAAACATGGCAACGTG 1912

Qy    1801 AAGGATATCGACAGCACTGATCATGACAGATGGTGTGAATACATTATGTATCGAGGGCTG 1860

Db    1913 AAGGATATTGACAGCTCAGATGGAGACAGATGGTGTGAGTACATCATGTACCGGGGTCTG 1972

Qy    1861 ATCAGGTTGGGTTATGCAAGAATCTCCCATGCTGAACTGAGTGATTCAGAAATTCAGATG 1920

Db    1973 ATCAGGTTGGGCTATGCAAGAATCTCTCATGCAGAACTGAGTGACTCTGAGATTCAGATG 2032

Qy    1921 GCAAAATTTAGGATCCCTGATGACCCCACTAATTATAGAGACAACCAGAAAGTGGTCATA 1980

Db    2033 GCCAAATTCAGGATCCCGGATGATCCCGCCAATTACAGAGACAACCAGAAAGTCGTCATA 2092

Qy    1981 GACCACAGAGAAGTTTCTGAGAAAATTCATTTTAATCCCAGATTTGGATCCTACAAAGAA 2040

Db    2093 GACCACCGAGGAGTTCCCAAGAACATTCACTTCAACCCCAGATTTGGCTCCTACAAAGAA 2152

Qy    2041 GGACACAATTATGAAAACAACCATCATTTTCATATGAATACTCCCAAATACTTTTTATGA 2100

Db    2153 GGACACAATTATGAGAACACCCATCATTTTCACATGAATACACCCAAATACTTCGTTTGA 2212

Qy    2101 AAC 2103

Db    2213 ACC 2215
```

RESULT 3
AC020593

```
LOCUS       AC020593              184206 bp    DNA     linear   PRI 07-NOV-2001
DEFINITION  Homo sapiens BAC clone RP11-317G22 from 4, complete sequence.
ACCESSION   AC020593
VERSION     AC020593.6  GI:14589709
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 184206)
  AUTHORS   Sulston,J.E. and Waterston,R.
  TITLE     Toward a complete human genome sequence
  JOURNAL   Genome Res. 8 (11), 1097-1108 (1998)
  MEDLINE   99063792
REFERENCE   2  (bases 1 to 184206)
  AUTHORS   Swearengen,S. and Cotton,M.
  TITLE     The sequence of Homo sapiens BAC clone RP11-317G22
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 184206)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-JAN-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   4  (bases 1 to 184206)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-JUL-2001) Genome Sequencing Center, Washington
```

University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
REFERENCE 5 (bases 1 to 184206)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (07-NOV-2001) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT On Jul 3, 2001 this sequence version replaced gi:12057004.
---------- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: sapiens@watson.wustl.edu
---------- Summary Statistics
Center project name: H_NH0317G22
----------.

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see http://genome.wustl.edu/gsc

SOURCE INFORMATION:
The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E., Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (http://bacpac.med.buffalo.edu)
VECTOR: pBACe3.6
NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the left is AC011956. Actual start of this clone is at base position 1 of RP11-317G22; actual end is at base position 184206 of RP11-317G22.

The sequence fidelity of RP11-317G22 between bases 17725 to 17956 can not be guaranteed due to an unresolved dinucleotide repeat. The sequence fidelity of RP11-317G22 between bases 86965 to 87015 can not be guaranteed due to an unresolved homopolymeric run.

FEATURES Location/Qualifiers
source 1. .184206
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="4"
/map="4"
/clone="RP11-317G22"
/clone_lib="RPCI-11"
repeat_region 5. .70
/rpt_family="L2"
misc_feature 252. .374
/note="match to EST BG169511 (NID:g12676214)"
misc_feature 252. .374
/note="match to EST AI088656 (NID:g3427715) qb14a11.x1"
misc_feature 252. .374
/note="match to EST BG214746 (NID:g13740767)"
misc_feature 252. .374
/note="match to EST BF055447 (NID:g10809343)"
misc_feature 252. .374
/note="match to EST AI077951 (NID:g3412359) oy39a10.s1"
misc_feature 252. .374
/note="match to EST AW006753 (NID:g5855531) wt07a02.x1"
misc_feature 252. .374
/note="match to EST BG291592 (NID:g13049843)"
misc_feature 252. .374
/note="match to EST W80356 (NID:g1391393) zh49a05.s1"
misc_feature 252. .494
/note="match to EST AA464688 (NID:g2189572) zx82a02.s1"
misc_feature 252. .374
/note="match to EST AW005009 (NID:g5853713) wv02e08.x1"
misc_feature 252. .374
/note="match to EST BG165990 (NID:g12672693)"
misc_feature 256. .374
/note="similar to Bos taurus EST AV609860 (NID:g9745530)"
misc_feature 280. .374
/note="match to EST AW371794 (NID:g6876448)"
misc_feature 347. .374
/note="match to EST AA303948 (NID:g1956281)"
misc_feature 708. .1203
/note="match to EST BI063872 (NID:g14471399)"
misc_feature 1021. .1308
/note="match to EST T06930 (NID:g318079)"
misc_feature 1119. .1143
/note="match to EST W68728 (NID:g1377608) zd37c01.s1"
misc_feature 1119. .1143
/note="match to EST AI220135 (NID:g3802338) qi44g05.x1"
misc_feature 1119. .1143
/note="match to EST AI271747 (NID:g3890914) qj77b10.x1"
misc_feature 1119. .1143
/note="match to EST BG167015 (NID:g12673728)"
misc_feature 1183. .1538
/note="match to EST BG546183 (NID:g13544848)"
repeat_region 1319. .1350
/rpt_family="AT_rich"
repeat_region 1349. .1466
/rpt_family="Alu"
misc_feature 1490. .1538
/note="match to EST AA847380 (NID:g2933898) of07d05.s1"
misc_feature 1490. .1538
/note="match to EST AI220135 (NID:g3802338) qi44g05.x1"
misc_feature 1490. .1538
/note="match to EST W68728 (NID:g1377608) zd37c01.s1"
misc_feature 1490. .1538
/note="match to EST BG167015 (NID:g12673728)"
misc_feature 1490. .1538
/note="match to EST AI271747 (NID:g3890914) qj77b10.x1"
misc_feature 1491. .1538
/note="match to EST W80356 (NID:g1391393) zh49a05.s1"
misc_feature 1491. .1538
/note="match to EST AI088656 (NID:g3427715) qb14a11.x1"
misc_feature 1491. .1508
/note="match to EST BG291592 (NID:g13049843)"
misc_feature 1491. .1538
/note="similar to Homo sapiens EST BG207366 (NID:g13729053)"
misc_feature 1491. .1538
/note="match to EST BG165990 (NID:g12672693)"
misc_feature 1491. .1538
/note="match to EST AA303948 (NID:g1956281)"
misc_feature 1491. .1538
/note="match to EST AW371794 (NID:g6876448)"
misc_feature 1491. .1538
/note="match to EST AW005009 (NID:g5853713) wv02e08.x1"
misc_feature 1491. .1538
/note="similar to Bos taurus EST AV609860 (NID:g9745530)"
misc_feature 1491. .1538
/note="match to EST AI077951 (NID:g3412359) oy39a10.s1"
misc_feature 1491. .1538
/note="match to EST BG214746 (NID:g13740767)"
misc_feature 1491. .1538
/note="match to EST AW006753 (NID:g5855531) wt07a02.x1"

```
     misc_feature    1491. .1538
                     /note="match to EST BG169511 (NID:g12676214)"
     misc_feature    1491. .1546
                     /note="match to EST BF055447 (NID:g10809343)"
     repeat_region   1561. .1606
                     /rpt_family="AT_rich"
     repeat_region   1696. .1986
                     /rpt_family="Alu"
     repeat_region   2175. .2217
                     /rpt_family="L1"
     repeat_region   2311. .2780
                     /rpt_family="MaLR"
     repeat_region   2828. .3176
                     /rpt_family="MaLR"
     repeat_region   3217. .3544
                     /rpt_family="Alu"
     repeat_region   3617. .3648
                     /rpt_family="AT_rich"
     repeat_region   3793. .4078
                     /rpt_family="Alu"
     repeat_region   4138. .4267
                     /rpt_family="Alu"
     repeat_region   4462. .4655
                     /rpt_family="MER1_type"
     misc_feature    4639. .4688
                     /note="match to EST AW005009 (NID:g5853713) wv02e08.x1"

  Query Match             12.7%;  Score 267.4;  DB 9;  Length 184206;
  Best Local Similarity   94.5%;  Pred. No. 7.6e-64;
  Matches  277;  Conservative    0;  Mismatches   16;  Indels    0;  Gaps    0;

Qy     799 ACAGGAACAGCTTCAGCTGCGGGGCTCCTTTACCTGCACACATGGGCAGCTGCTGTGTCT 858
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 111216 ACAGGAACAGCTTCAGCTGCGGGGCTCCTTTACCTGCACACATGGGCAGCTGCTGTGTCT 111275

Qy     859 GGCTGTGTCTTCGCCATCTTTACTGCATCCATGTGGCCCCAAACACTTGGACACCTTATT 918
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 111276 GGCTGTGTCTTCGCCATCTTTACTGCATCCATGTGGCCCCAAACACTTGGACACCTTATT 111335

Qy     919 AACTCAGGGACAAACCCTGGGAAAACCATGACCATTGCCATGATATTTTATCTTCTAGAA 978
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 111336 AACTCAGGGACAAACCCTGGGAAAACCATGACCATTGCCATGATATTTTATCTTCTAGAA 111395

Qy     979 ATATTTTTCTGTGCCTGGTGCACAGCTTTTAAGTTTGTCCCAGGAGGTGTCTACGCTAGA 1038
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 111396 ATATTTTTCTGTGCCTGGTGCACAGCTTTTAAGTTTGTCCCAGGAGGTGTCTACGCTAGA 111455

Qy    1039 GAAAGATCAGATGTGCTTTTGGGGACAATGATGTTAATTATCGGGCTGAATAT 1091
           ||||||||||||||||||||||  || | ||  || || | || ||
Db 111456 GAAAGATCAGATGTGCTTTTGGGTGAGTACATTTGAAAGGTCTGGATTAAAAT 111508


RESULT 4
AL158816
LOCUS       AL158816               222885 bp    DNA     linear   HTG 10-JUL-2001
DEFINITION  Homo sapiens chromosome 6 clone RP3-405B4, *** SEQUENCING IN
            PROGRESS ***, 33 unordered pieces.
ACCESSION   AL158816
VERSION     AL158816.11  GI:9943990
KEYWORDS    HTG; HTGS_PHASE1; HTGS_CANCELLED.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 222885)
  AUTHORS   Sims,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Aug 29, 2000 this sequence version replaced gi:9926474.
            -------------- Genome Center
            Center: Sanger Centre
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            -------------- Project Information
            Center project name: dJ405B4
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Sequencing vector: plasmid; L08752; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Consensus quality: 204390 bases at least Q40
            Consensus quality: 211527 bases at least Q30
            Consensus quality: 215491 bases at least Q20
            Insert size: 219685; sum-of-contigs
            Insert size: 126549; 14.0% error; agarose-fp
            Quality coverage: 3.63x in Q20 bases; sum-of-contigs Quality
            coverage: 6.72x in Q20 bases; agarose-fp
            --------------
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 33 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1     3599: contig of 3599 bp in length
            *    3600 3699: gap of      100 bp
            *    3700    12818: contig of 9119 bp in length
            *   12819 12918: gap of      100 bp
            *   12919    16398: contig of 3480 bp in length
            *   16399 16498: gap of      100 bp
            *   16499    18590: contig of 2092 bp in length
            *   18591 18690: gap of      100 bp
            *   18691    22372: contig of 3682 bp in length
            *   22373 22472: gap of      100 bp
            *   22473    26905: contig of 4433 bp in length
            *   26906 27005: gap of      100 bp
            *   27006    29581: contig of 2576 bp in length
            *   29582 29681: gap of      100 bp
            *   29682    33446: contig of 3765 bp in length
            *   33447 33546: gap of      100 bp
            *   33547    48536: contig of 14990 bp in length
            *   48537 48636: gap of      100 bp
            *   48637    55599: contig of 6963 bp in length
            *   55600 55699: gap of      100 bp
            *   55700    65124: contig of 9425 bp in length
            *   65125 65224: gap of      100 bp
            *   65225    69090: contig of 3866 bp in length
            *   69091 69190: gap of      100 bp
            *   69191    73894: contig of 4704 bp in length
            *   73895 73994: gap of      100 bp
            *   73995    82665: contig of 8671 bp in length
            *   82666 82765: gap of      100 bp
            *   82766    87239: contig of 4474 bp in length
            *   87240 87339: gap of      100 bp
            *   87340    93671: contig of 6332 bp in length
            *   93672 93771: gap of      100 bp
            *   93772    97550: contig of 3779 bp in length
            *   97551 97650: gap of      100 bp
            *   97651   100101: contig of 2451 bp in length
            *  100102 100201: gap of      100 bp
            *  100202   103416: contig of 3215 bp in length
            *  103417 103516: gap of      100 bp
            *  103517   107564: contig of 4048 bp in length
            *  107565 107664: gap of      100 bp
            *  107665   131778: contig of 24114 bp in length
            *  131779 131878: gap of      100 bp
            *  131879   135261: contig of 3383 bp in length
            *  135262 135361: gap of      100 bp
            *  135362   137913: contig of 2552 bp in length
            *  137914 138013: gap of      100 bp
            *  138014   140949: contig of 2936 bp in length
            *  140950 141049: gap of      100 bp
```

```
            *   141050   156466: contig of 15417 bp in length
            *   156467 156566: gap of      100 bp
            *   156567   164836: contig of 8270 bp in length
            *   164837 164936: gap of      100 bp
            *   164937   168950: contig of 4014 bp in length
            *   168951 169050: gap of      100 bp
            *   169051   173724: contig of 4674 bp in length
            *   173725 173824: gap of      100 bp
            *   173825   179991: contig of 6167 bp in length
            *   179992 180091: gap of      100 bp
            *   180092   190208: contig of 10117 bp in length
            *   190209 190308: gap of      100 bp
            *   190309   193471: contig of 3163 bp in length
            *   193472 193571: gap of      100 bp
            *   193572   197829: contig of 4258 bp in length
            *   197830 197929: gap of      100 bp
            *   197930   222885: contig of 24956 bp in length.
FEATURES            Location/Qualifiers
     source         1. .222885
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /chromosome="6"
                    /clone="RP3-405B4"
                    /clone_lib="RPCI-3"
     misc_feature   1. .3599
                    /note="assembly_fragment:00027"
     misc_feature   3700. .12818
                    /note="assembly_fragment:00056"
     misc_feature   12919. .16398
                    /note="assembly_fragment:00093"
     misc_feature   16499. .18590
                    /note="assembly_fragment:00125"
     misc_feature   18691. .22372
                    /note="assembly_fragment:00140"
     misc_feature   22473. .26905
                    /note="assembly_fragment:00167"
     misc_feature   27006. .29581
                    /note="assembly_fragment:00175"
     misc_feature   29682. .33446
                    /note="assembly_fragment:00178"
     misc_feature   33547. .48536
                    /note="assembly_fragment:00416"
     misc_feature   48637. .55599
                    /note="assembly_fragment:00542"
     misc_feature   55700. .65124
                    /note="assembly_fragment:00565"
     misc_feature   65225. .69090
                    /note="assembly_fragment:00583"
     misc_feature   69191. .73894
                    /note="assembly_fragment:00649"
     misc_feature   73995. .82665
                    /note="assembly_fragment:00709"
     misc_feature   82766. .87239
                    /note="assembly_fragment:00789"
     misc_feature   87340. .93671
                    /note="assembly_fragment:00874"
     misc_feature   93772. .97550
                    /note="assembly_fragment:00924"
     misc_feature   97651. .100101
                    /note="assembly_fragment:01052"
     misc_feature   100202. .103416
                    /note="assembly_fragment:01082"
     misc_feature   103517. .107564
                    /note="assembly_fragment:01216"
     misc_feature   107665. .131778
                    /note="assembly_fragment:01234"
     misc_feature   131879. .135261
                    /note="assembly_fragment:01280"
     misc_feature   135362. .137913
                    /note="assembly_fragment:01337"
     misc_feature   138014. .140949
                    /note="assembly_fragment:01643"
     misc_feature   141050. .156466
                    /note="assembly_fragment:01660"
     misc_feature   156567. .164836
                    /note="assembly_fragment:01690"
     misc_feature   164937. .168950
                    /note="assembly_fragment:01692.0"
     misc_feature   169051. .173724
                    /note="assembly_fragment:01748"
     misc_feature   173825. .179991
                    /note="assembly_fragment:02061"
     misc_feature   180092. .190208
                    /note="assembly_fragment:02140"
     misc_feature   190309. .193471
                    /note="assembly_fragment:02150"
     misc_feature   193572. .197829
                    /note="assembly_fragment:02181"
     misc_feature   197930. .222885
                    /note="assembly_fragment:00834
                    clone_end:T7
                    vector_side:right"
BASE COUNT    69666 a  41364 c  44947 g  63682 t   3226 others
ORIGIN

  Query Match             7.6%;  Score 160;  DB 2;  Length 222885;
  Best Local Similarity   89.6%;  Pred. No. 1.4e-33;
  Matches  172;  Conservative    0;  Mismatches   20;  Indels    0;  Gaps    0;

Qy    1863 CAGGTTGGGTTATGCAAGAATCTCCCATGCTGAACTGAGTGATTCAGAAATTCAGATGGC 1922
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  157674 CAGGTTGGGTTATGCAAGAATCTCCCATGCTGAACTGAGTGATTCAGAAATTCAGATGGC 157733

Qy    1923 AAAATTTAGGATCCCTGATGACCCCACTAATTATAGAGACAACCAGAAAGTGGTCATAGA 1982
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  157734 AAAATTTAGGATCCCTGATGACCCCACTAATTATAGAGACAACCAGAAAGTGGTCATAGA 157793

Qy    1983 CCACAGAGAAGTTTCTGAGAAAATTCATTTTAATCCCAGATTTGGATCCTACAAAGAAGG 2042
           |||||||||||||||||||||||||||||||||||| |  |   | |       |
Db  157794 CCACAGAGAAGTTTCTGAGAAAATTCATTTTAATCCCAGGTGAGTTCCTTTATGCTGTAG 157853

Qy    2043 ACACAATTATGA 2054
             || ||||||
Db  157854 TTCCAGTTATGA 157865


RESULT 5
AL158816/c
LOCUS       AL158816              222885 bp    DNA     linear   HTG 10-JUL-2001
DEFINITION  Homo sapiens chromosome 6 clone RP3-405B4, *** SEQUENCING IN
            PROGRESS ***, 33 unordered pieces.
ACCESSION   AL158816
VERSION     AL158816.11  GI:9943990
KEYWORDS    HTG; HTGS_PHASE1; HTGS_CANCELLED.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 222885)
  AUTHORS   Sims,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Aug 29, 2000 this sequence version replaced gi:9926474.
            -------------- Genome Center
            Center: Sanger Centre
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            -------------- Project Information
            Center project name: dJ405B4
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Sequencing vector: plasmid; L08752; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
```

```
Consensus quality: 204390 bases at least Q40
Consensus quality: 211527 bases at least Q30
Consensus quality: 215491 bases at least Q20
Insert size: 219685; sum-of-contigs
Insert size: 126549; 14.0% error; agarose-fp
Quality coverage: 3.63x in Q20 bases; sum-of-contigs Quality
coverage: 6.72x in Q20 bases; agarose-fp
--------------
* NOTE: This is a 'working draft' sequence. It currently
* consists of 33 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*        1     3599: contig of 3599 bp in length
*     3600 3699: gap of      100 bp
*     3700    12818: contig of 9119 bp in length
*    12819 12918: gap of      100 bp
*    12919    16398: contig of 3480 bp in length
*    16399 16498: gap of      100 bp
*    16499    18590: contig of 2092 bp in length
*    18591 18690: gap of      100 bp
*    18691    22372: contig of 3682 bp in length
*    22373 22472: gap of      100 bp
*    22473    26905: contig of 4433 bp in length
*    26906 27005: gap of      100 bp
*    27006    29581: contig of 2576 bp in length
*    29582 29681: gap of      100 bp
*    29682    33446: contig of 3765 bp in length
*    33447 33546: gap of      100 bp
*    33547    48536: contig of 14990 bp in length
*    48537 48636: gap of      100 bp
*    48637    55599: contig of 6963 bp in length
*    55600 55699: gap of      100 bp
*    55700    65124: contig of 9425 bp in length
*    65125 65224: gap of      100 bp
*    65225    69090: contig of 3866 bp in length
*    69091 69190: gap of      100 bp
*    69191    73894: contig of 4704 bp in length
*    73895 73994: gap of      100 bp
*    73995    82665: contig of 8671 bp in length
*    82666 82765: gap of      100 bp
*    82766    87239: contig of 4474 bp in length
*    87240 87339: gap of      100 bp
*    87340    93671: contig of 6332 bp in length
*    93672 93771: gap of      100 bp
*    93772    97550: contig of 3779 bp in length
*    97551 97650: gap of      100 bp
*    97651   100101: contig of 2451 bp in length
*   100102 100201: gap of      100 bp
*   100202   103416: contig of 3215 bp in length
*   103417 103516: gap of      100 bp
*   103517   107564: contig of 4048 bp in length
*   107565 107664: gap of      100 bp
*   107665   131778: contig of 24114 bp in length
*   131779 131878: gap of      100 bp
*   131879   135261: contig of 3383 bp in length
*   135262 135361: gap of      100 bp
*   135362   137913: contig of 2552 bp in length
*   137914 138013: gap of      100 bp
*   138014   140949: contig of 2936 bp in length
*   140950 141049: gap of      100 bp
*   141050   156466: contig of 15417 bp in length
*   156467 156566: gap of      100 bp
*   156567   164836: contig of 8270 bp in length
*   164837 164936: gap of      100 bp
*   164937   168950: contig of 4014 bp in length
*   168951 169050: gap of      100 bp
*   169051   173724: contig of 4674 bp in length
*   173725 173824: gap of      100 bp
*   173825   179991: contig of 6167 bp in length
*   179992 180091: gap of      100 bp
*   180092   190208: contig of 10117 bp in length
*   190209 190308: gap of      100 bp
*   190309   193471: contig of 3163 bp in length
*   193472 193571: gap of      100 bp
*   193572   197829: contig of 4258 bp in length
*   197830 197929: gap of      100 bp
*   197930   222885: contig of 24956 bp in length.
FEATURES             Location/Qualifiers
     source          1. .222885
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /clone="RP3-405B4"
                     /clone_lib="RPCI-3"
     misc_feature    1. .3599
                     /note="assembly_fragment:00027"
     misc_feature    3700. .12818
                     /note="assembly_fragment:00056"
     misc_feature    12919. .16398
                     /note="assembly_fragment:00093"
     misc_feature    16499. .18590
                     /note="assembly_fragment:00125"
     misc_feature    18691. .22372
                     /note="assembly_fragment:00140"
     misc_feature    22473. .26905
                     /note="assembly_fragment:00167"
     misc_feature    27006. .29581
                     /note="assembly_fragment:00175"
     misc_feature    29682. .33446
                     /note="assembly_fragment:00178"
     misc_feature    33547. .48536
                     /note="assembly_fragment:00416"
     misc_feature    48637. .55599
                     /note="assembly_fragment:00542"
     misc_feature    55700. .65124
                     /note="assembly_fragment:00565"
     misc_feature    65225. .69090
                     /note="assembly_fragment:00583"
     misc_feature    69191. .73894
                     /note="assembly_fragment:00649"
     misc_feature    73995. .82665
                     /note="assembly_fragment:00709"
     misc_feature    82766. .87239
                     /note="assembly_fragment:00789"
     misc_feature    87340. .93671
                     /note="assembly_fragment:00874"
     misc_feature    93772. .97550
                     /note="assembly_fragment:00924"
     misc_feature    97651. .100101
                     /note="assembly_fragment:01052"
     misc_feature    100202. .103416
                     /note="assembly_fragment:01082"
     misc_feature    103517. .107564
                     /note="assembly_fragment:01216"
     misc_feature    107665. .131778
                     /note="assembly_fragment:01234"
     misc_feature    131879. .135261
                     /note="assembly_fragment:01280"
     misc_feature    135362. .137913
                     /note="assembly_fragment:01337"
     misc_feature    138014. .140949
                     /note="assembly_fragment:01643"
     misc_feature    141050. .156466
                     /note="assembly_fragment:01660"
     misc_feature    156567. .164836
                     /note="assembly_fragment:01690"
     misc_feature    164937. .168950
                     /note="assembly_fragment:01692.0"
     misc_feature    169051. .173724
                     /note="assembly_fragment:01748"
     misc_feature    173825. .179991
                     /note="assembly_fragment:02061"
```

```
     misc_feature    180092. .190208
                     /note="assembly_fragment:02140"
     misc_feature    190309. .193471
                     /note="assembly_fragment:02150"
     misc_feature    193572. .197829
                     /note="assembly_fragment:02181"
     misc_feature    197930. .222885
                     /note="assembly_fragment:00834
                     clone_end:T7
                     vector_side:right"
BASE COUNT    69666 a  41364 c  44947 g  63682 t   3226 others
ORIGIN

  Query Match             7.2%;  Score 151;  DB 2;  Length 222885;
  Best Local Similarity   100.0%;  Pred. No. 5e-31;
  Matches  151;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    1508 GGATTATGGCTTTGTCAAGATACCCAATTGTGAAATCTGAGCATCACCTTCTTCCGTCAC 1567
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 215729 GGATTATGGCTTTGTCAAGATACCCAATTGTGAAATCTGAGCATCACCTTCTTCCGTCAC 215670

Qy    1568 CAGAGGGCGAGATCGCACCAGCCATCACATTGACCGTTAACATTTCGGGCAAGCTGGTGG 1627
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 215669 CAGAGGGCGAGATCGCACCAGCCATCACATTGACCGTTAACATTTCGGGCAAGCTGGTGG 215610

Qy    1628 ATTTTGTCGTGACACACTTTGGGAACCACGA 1658
           ||||||||||||||||||||||||||||||
Db 215609 ATTTTGTCGTGACACACTTTGGGAACCACGA 215579


RESULT 6
SCCHRIII/c
LOCUS       SCCHRIII              316613 bp    DNA     linear   PLN 29-JUN-2001
DEFINITION  S.cerevisiae chromosome III complete DNA sequence.
ACCESSION   X59720 S43845 S49180 S58084 S93798
VERSION     X59720.2  GI:14588895
KEYWORDS    chromosome.
SOURCE      baker's yeast.
  ORGANISM  Saccharomyces cerevisiae
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE   1  (bases 1 to 316613)
  AUTHORS   Rad,M.R., Lutzenkirchen,K., Xu,G., Kleinhans,U. and Hollenberg,C.P.
  TITLE     The complete sequence of a 11,953 bp fragment from C1G on
            chromosome III encompasses four new open reading frames
  JOURNAL   Yeast 7 (5), 533-538 (1991)
  MEDLINE   91377317
REFERENCE   2  (bases 1 to 316613)
  AUTHORS   Wilson,C., Bergantino,E., Lanfranchi,G., Valle,G., Carignani,G. and
            Frontali,L.
  TITLE     A putative serine/threonine protein kinase gene on chromosome III
            of Saccharomyces cerevisiae
  JOURNAL   Yeast 8 (1), 71-77 (1992)
  MEDLINE   92254506
REFERENCE   3  (bases 1 to 316613)
  AUTHORS   Biteau,N., Fremaux,C., Hebrard,S., Menara,A., Aigle,M. and
            Crouzet,M.
  TITLE     The complete sequence of a 10.8kb fragment to the right of the
            chromosome III centromere of Saccharomyces cerevisiae
  JOURNAL   Yeast 8 (1), 61-70 (1992)
  MEDLINE   92254505
REFERENCE   4  (bases 1 to 316613)
  AUTHORS   Benit,P., Chanet,R., Fabre,F., Faye,G., Fukuhara,H. and Sor,F.
  TITLE     Sequence of the sup61-RAD18 region on chromosome III of
            Saccharomyces cerevisiae
  JOURNAL   Yeast 8 (2), 147-153 (1992)
  MEDLINE   92221691
REFERENCE   5  (bases 1 to 316613)
  AUTHORS   Bolle,P.A., Gilliquet,V., Berben,G., Dumont,J. and Hilger,F.
  TITLE     The complete sequence of K3B, a 7.9 kb fragment between PGK1 and
            CRY1 on chromosome III, reveals the presence of seven open reading
            frames
  JOURNAL   Yeast 8 (3), 205-213 (1992)
  MEDLINE   92245758
REFERENCE   6  (bases 1 to 316613)
  AUTHORS   Sor,F., Cheret,G., Fabre,F., Faye,G. and Fukuhara,H.
  TITLE     Sequence of the HMR region on chromosome III of Saccharomyces
            cerevisiae
  JOURNAL   Yeast 8 (3), 215-222 (1992)
  MEDLINE   92245759
REFERENCE   7  (bases 1 to 316613)
  AUTHORS   Skala,J., Purnelle,B. and Goffeau,A.
  TITLE     The complete sequence of a 10.8 kb segment distal of SUF2 on the
            right arm of chromosome III from Saccharomyces cerevisiae reveals
            seven open reading frames including the RVS161, ADP1 and PGK genes
  JOURNAL   Yeast 8 (5), 409-417 (1992)
  MEDLINE   92327849
REFERENCE   8  (bases 1 to 316613)
  AUTHORS   Oliver,S.G., van der Aart,Q.J., Agostoni-Carbone,M.L., Aigle,M.,
            Alberghina,L., Alexandraki,D., Antoine,G., Anwar,R., Ballesta,J.P.,
            Benit,P. et al.
  TITLE     The complete DNA sequence of yeast chromosome III
  JOURNAL   Nature 357 (6373), 38-46 (1992)
  MEDLINE   92244356
REFERENCE   9  (bases 1 to 316613)
  AUTHORS   Scherens,B., Messenguy,F., Gigot,D. and Dubois,E.
  TITLE     The complete sequence of a 9,543 bp segment on the left arm of
            chromosome III reveals five open reading frames including
            glucokinase and the protein disulfide isomerase
  JOURNAL   Yeast 8 (7), 577-585 (1992)
  MEDLINE   92397595
REFERENCE   10 (bases 1 to 316613)
  AUTHORS   Wilson,C., Grisanti,P. and Frontali,L.
  TITLE     The complete sequence of a 6146 bp fragment of Saccharomyces
            cerevisiae chromosome III contains two new open reading frames
  JOURNAL   Yeast 8 (7), 569-575 (1992)
  MEDLINE   92397594
REFERENCE   11 (bases 26740 to 32076)
  AUTHORS   Defoor,E., Debrabandere,R., Keyers,B., Voet,M. and Volckaert,G.
  TITLE     Nucleotide sequence of D10B, a BamHI fragment on the small-ring
            chromosome III of Saccharomyces cerevisiae
  JOURNAL   Yeast 8 (8), 681-687 (1992)
  MEDLINE   93070606
REFERENCE   12 (bases 169581 to 171116; 171683 to 172169)
  AUTHORS   Agostoni Carbone,M.L., Panzeri,L., Muzi Falconi,M., Carcano,C.,
            Plevani,P. and Lucchini,G.
  TITLE     Nucleotide sequence of 9.2 kb left of CRY1 on yeast chromosome III
            from strain AB972: evidence for a Ty insertion and functional
            analysis of open reading frame YCR28
  JOURNAL   Yeast 8 (9), 805-812 (1992)
  MEDLINE   93070619
REFERENCE   13 (bases 1 to 315338)
  AUTHORS   Valle,G.
  TITLE     TA-repeat microsatellites are closely associated with ARS consensus
            sequences in yeast chromosome III
  JOURNAL   Yeast 9 (7), 753-759 (1993)
  MEDLINE   93377412
REFERENCE   14 (bases 1 to 315338)
  AUTHORS   Slonimski,P.P. and Brouillet,S.
  TITLE     A data-base of chromosome III of Saccharomyces cerevisiae
  JOURNAL   Yeast 9 (9), 941-1029 (1993)
  MEDLINE   94091061
REFERENCE   15 (bases 1 to 315339)
  AUTHORS   Rodriguez-Cousino,N., Lill,R., Neupert,W. and Court,D.A.
  TITLE     Identification and initial characterization of the cytosolic
            protein Ycr77p
  JOURNAL   Yeast 11 (6), 581-585 (1995)
  MEDLINE   95373282
REFERENCE   16 (bases 1 to 316613)
  AUTHORS   MIPS.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-MAR-1992) MIPS, D-8033 Martinsried, FRG. Data
            collected by MIPS on behalf of the European Yeast Chromosome III
            Sequencing project
  REMARK    replaced by [13]
REFERENCE   17 (bases 1 to 314957)
```

AUTHORS Jimenez,A.
TITLE Direct Submission
JOURNAL Submitted (28-DEC-1992) MIPS, D-8033 Martinsried, FRG. Data collected by MIPS on behalf of the European Yeast Chromosome III Sequencing project. Update originating from A. Jimenez
REMARK replaced by [14]
REFERENCE 18 (bases 1 to 316613)
AUTHORS Louis,E.J.
TITLE Direct Submission
JOURNAL Submitted (25-JUN-1993) MIPS, D-8033 Martinsried, FRG. Data collected by MIPS on behalf of the European Yeast Chromosome III Sequencing project. Update originating from E.J. Louis
REMARK revised by [18]
REFERENCE 19 (bases 1 to 316613)
AUTHORS Louis,E.J.
TITLE Direct Submission
JOURNAL Submitted (14-FEB-1995) MIPS, D-8033 Martinsried, FRG. Data collected by MIPS on behalf of the European Yeast Chromosome III Sequencing project. Update originating from E.J. Louis
REMARK Revised by [20]
REFERENCE 20 (bases 1 to 316613)
AUTHORS Gromadka,R.
TITLE Direct Submission
JOURNAL Submitted (29-JAN-1996) R. Gromadka, Protein Biosynthesis, Institute of Biochemistry and Biophysics, Pawinskiego 5A, Warsaw, Poland, Electronic Mail Address: robert@psd.ibb.waw.pl
REMARK revised by [21]
REFERENCE 21 (bases 1 to 316613)
AUTHORS MIPS.
TITLE Direct Submission
JOURNAL Submitted (17-JUN-2001) MIPS Yeast Genome Database, GSF - Ingolstaedter Landstrasse 1 D-85764 Neuherberg, Germany, mips-yeast-adm@gsf.de
REMARK Resequencing project, achieved by the joint effort of G. Valles and G. Volckaerts laboratories. Munich information center for protein sequences.
COMMENT On Jul 3, 2001 this sequence version replaced gi:1907116.
The contig sequence which has been released to the file server in 1992 has been subject to a resequencing project, achieved by the joint effort of G. Valles and G. Volckaerts laboratories. The following lines present a summary of the altered entities. The resequenced chrIII contig is 316613 bp with GCG-check: 429 GENETIC ENTITIES valid no longer:
YCL012w (YCL014w + YCL012w -> YCL014w / C_B1367 + C_C231->C_A1636), PEP Y, ORF in identical form now C-terminal part of YCL014w. (A frame shift in
the nt sequence of YCL014w leads to an elongated polypeptide.) This elongated polypeptide encoded by the current YCL014w resembles the Bud3p polypeptide from the
EMBL-Entry (Accession:SC175801;ID:U17580, total length 1636 aa)
YCL006c, PEP Y, ORF obsolete, is not an ORF in current Chromosome III contig.
YCR062w (YCR061w + YCR062w -> YCR061w / C_A583 + C_B120 -> C_B631), PEP Y, ORF in identical form now C-terminal part of YCR061w.
YCR068w-a (YCR068w + YCR068w-a -> YCR068w / C_C429 + C_A145 -> C_A520), PEP Y, ORF in identical form now C-terminal part of YCR068w.
YCR103c (C_F111, former ORF in region corresponding to coord. 307255-307587(C)) is obsolete, open reading frame is destroyed by single nt exchange resulting in a
stop codon after 45 aa.
NEW GENETIC ENTITIES:
YCL26c-b (C_E193(YCL027c-a)), PEP Y, new ORF
YCL021w-a (C_C125), PEP Y, new orf
ALTERED GENETIC ENTITIES:
YCL076w, PEP Y, peptide elongated N-terminally by 51 aa YCL074w, PEP Y, 3 aa exchanged
YCL073c, PEP Y, 4 aa exchanged
YCL068c (C_F190->C_E260), PEP Y, peptide elongated N-terminally by 70 aa YCL065w, PEP Y, 1 aa exchanged
YCL064c, PEP Y, 3 aa exchanged
YCL061c (C_D853->C_F1096), PEP Y, peptide elongated N-terminally by 243 aa YCL051w, PEP Y, 11 aa altered, 6 aa exchanged, 1 aa additional, 4 aa less at the C-terminus
YCL050c, PEP Y, 1 aa exchanged
YCL049c, PEP Y, 4 aa exchanged
YCL042w, PEP Y, 8 aa altered, 7 aa exchanged, 1 aa additional
YCL034w (C_A265->C_B354), PEP Y, peptide elongated N-terminally by 89 aa YCL030c, PEP Y, 1 aa exchanged
YCL028w, PEP Y, 1 aa exchanged
YCL027w, PEP Y, 3 aa exchanged
YCL026c-a, PEP Y, 8 aa altered, 7 aa exchanged, 1 aa additional
YCL025c (C_F633->C_F595), PEP Y, 7 aa exchanged, the c-terminus is shorter by 38 aa and the last 5 aa are altered
LTR deltaCL05, 1 nt exchange
LTR deltaCL04, 1 nt deletion
TY2A_C, PEP Y, 1 aa exchanged

Query Match 6.3%; Score 132.4; DB 8; Length 316613;
Best Local Similarity 52.3%; Pred. No. 9e-26;
Matches 342; Conservative 0; Mismatches 306; Indels 6; Gaps 2;

```
Qy   1281 GGTCTCTGCTGCCATCTGGCCTTTCAGGTTTGGATATGACAATGAAGGGTGGTCTAGTCT 1340
          | |  | ||||  || ||| || ||   |||||    || |||||   |||| |   |
Db 145589 GATTACAGCTGGTATTTGGACTATCCACTTTGGTCTCGATAATGATATGTGGGCATCTGA 145530

Qy   1341 AGAAAGATCAGCTCACCTGCTCAATGAAACAGGTGCAGATTTCATAACAATTTTGGAGAG 1400
          |||  |||        ||||  | || || |  |    |||| |  ||    |  | || |
Db 145529 AGACAGAATGATCAACCTTATTAAAGATATGGAACTAGATGTGGTAGGTCTACTAGAAAC 145470

Qy   1401 TGATGCTTCTAAGCCCTATATGGGGAACAATGACTTAACCATGTGGCTAGGGGAAAAGTT 1460
           |||  |     |          ||||||||||  ||  |||| |      ||||   |  | ||
Db 145469 AGATACACAAAGAATTACCATGGGGAACAGGGATCTAACTAGCAAACTAGCTCATGATTT 145410

Qy   1461 GGGTTTCTATACAGACTTTGGTCCAAGCACAAGGTATCACACTTGGGGGATTATGGCTTT 1520
          |  | | ||| |||| || || ||| |  |||    | || || |||||   | |   | |
Db 145409 GAATATGTATGCAGATTTCGGACCAGGTCCAAATAAACATACCTGGGGCTGTGTTCTTCT 145350

Qy   1521 GTCAAGATACCCAATTGTGAAATCTGAGCATCACCTTCTTCCGTCACCAGAGGGCGAGAT 1580
           || | || ||| || || || ||| ||||||  |  | || || |||| || ||  |
Db 145349 TTCTAAATTCCCTATCGTAAATTCTACGCATCATTTATTGCCCTCTCCAGTTGGGGAACT 145290

Qy   1581 CGCACCAGCCATCACATTGACCGTTAACATTTC---GGGCAAGCTGGTGGATTTTGTCGT 1637
           || ||||||||       ||  || | | |     | ||  || || ||  |  | ||
Db 145289 TGCGCCAGCCATTCATGCCACACTTCAAACGTACAATGACACTCTCGTTGACGTCTTTGT 145230

Qy   1638 GACACACTTTGGGAACCACGAAGATGACCTCGACAGGAAACTGCAGGCTATTGCTGTTTC 1697
            ||   |||  |  | || ||||     || || | ||||||   ||      |  |
Db 145229 ATTCCATAGTGGACAAGAAGAGGATGAAGAGGATAGAAGACTGCAAAGTAACTACATGGC 145170

Qy   1698 AAAACTACTGAAAAGTAGCTCTAATCAAGTGATATTTCTGGGATATATCACTTCAGCACC 1757
           || ||  ||   | ||   ||   | ||  || ||  |  | ||  |   |  | ||
Db 145169 TAAGCTCATGGGCAATACGACTCGCCCAGCTATTTTATTAAGTTACTTAGTTGTTGATCC 145110

Qy   1758 TGGCTCCAGAGATT---ATCTACAGCTCACTGAACATGGCAATGTGAAGGATATCGACAG 1814
           ||      | | |   ||    |  | | |||      |    || | || || ||  |
Db 145109 AGGTGAAGGCAACTACAATACGTACGTTAGTGAAACATCCGGAATGCACGACATTGATCC 145050

Qy   1815 CACTGATCATGACAGATGGTGTGAATACATTATGTATCGAGGGCTGATCAGGTTGGGTTA 1874
            ||||  |||| ||||||||||| || ||  ||||| | ||  |||  ||    || ||
Db 145049 TTCTGACGATGATAGATGGTGTGAGTATATCTTGTATAGGGGCTTGAGAAGAACAGGATA 144990

Qy   1875 TGCAAGAATCTCCCATGCTGAACTGAGTGATTCAGAAATTCAGATGGCAAAATT 1928
          ||| ||| |  |    |      | |  ||| | ||  | ||  | |  || ||
Db 144989 TGCTAGAGTTGCAAGAGGAACGATAACCGATACGGAGCTACAAGTTGGTAAGTT 144936
```

RESULT 7
D89130
LOCUS D89130 1234 bp mRNA linear PLN 13-MAR-1998
DEFINITION Schizosaccharomyces pombe mRNA, partial cds, clone: SY 0503.
ACCESSION D89130
VERSION D89130.1 GI:1749467
KEYWORDS .
SOURCE Schizosaccharomyces pombe (strain:PR745) cDNA to mRNA, clone_lib:library of H. Nojima clone:SY 0503.
ORGANISM Schizosaccharomyces pombe

```
            Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
            Schizosaccharomycetales; Schizosaccharomycetaceae;
            Schizosaccharomyces.
REFERENCE   1  (bases 1 to 1234)
  AUTHORS   Yoshioka,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-NOV-1996) Sachiyo Yoshioka, Tsukita Cell Axis Project
            ERATO JST, Kyoto Research Park; 17 Chudouji Minamimachi,
            Shimokyo-ku, Kyoto, Kyoto 600, Japan
            (E-mail:syoshi@cell.tsukita.jst.go.jp, Tel:+81-75-315-7913,
            Fax:+81-75-315-6420)
REFERENCE   2  (sites)
  AUTHORS   Yoshioka,S., Kato,K., Nakai,K., Okayama,H. and Nojima,H.
  TITLE     Identification of open reading frames in Schizosaccharomyces pombe
            cDNAs
  JOURNAL   DNA Res. 4 (6), 363-369 (1997)
  MEDLINE   98162722
  FEATURES             Location/Qualifiers
     source            1. .1234
                       /organism="Schizosaccharomyces pombe"
                       /strain="PR745"
                       /db_xref="taxon:4896"
                       /clone="SY 0503"
                       /clone_lib="library of H. Nojima"
     CDS               <1. .1207
                       /note="similar to Saccharomyce cerevisiae hypothetical
                       107.9KD protein in POL4-SRD1 intergenic region, SWISS-PROT
                       Accession Number P25618"
                       /codon_start=2
                       /protein_id="BAA13792.1"
                       /db_xref="GI:1749468"
                       /translation="IYSFASIRISSFYSPAKVWGGAFLVYILYSLAHVWVVAYEFVPG
                       GPILRERTSYILIFIGWNLAALVPAYSGESKEPNKADSSVVDIKQSDSSYRRRSFKKS
                       LLTGFCLALMALKFAIQNMPPYDYTPYHPNEKLFTAGIWTIHFGLDNFMYASENRIRD
                       AVRDMELDVFGLLESDTQRLIMGFRDLTQVLAHDLGMYADYGPGPDKHTWGAALLSKF
                       PIVNSTHHLLPSPQGELAPAIHATLDVYGELIDVVVSHNGQYESQLDRRLQSTELARI
                       MRESPRPLVFLGYVVSNVGQEPQTILTRDTGMLDIEPADYDRWCQYIFYRGVKRIGYA
                       RLHRSTITDTELQTGKFLVTKDLGRNVRIDKEHVPESHRYPSLFEGTGVNGHYYDNNL
                       VVHEPWYYD"
BASE COUNT      333 a    253 c    253 g    395 t
ORIGIN

  Query Match             5.8%;  Score 121.8;  DB 8;  Length 1234;
  Best Local Similarity   49.0%;  Pred. No. 5e-23;
  Matches  324;  Conservative    0;  Mismatches  337;  Indels    0;  Gaps    0;

Qy    1283 TCTCTGCTGCCATCTGGCCTTTCAGGTTTGGATATGACAATGAAGGGTGGTCTAGTCTAG 1342
           || |||||| || ||| || |    ||||| |  ||||||     ||   | |||
Db     408 TCACTGCTGGTATTTGGACTATTCACTTTGGCTTAGACAATTTCATGTATGCAAGTGAGA 467

Qy    1343 AAAGATCAGCTCACCTGCTCAATGAAACAGGTGCAGATTTCATAACAATTTTGGAGAGTG 1402
           |  |   |    |     |    || |  |     ||| || |     |  ||||   ||
Db     468 ATCGTATACGCGATGCTGTTCGAGATATGGAACTTGATGTCTTTGGTCTACTGGAATCTG 527

Qy    1403 ATGCTTCTAAGCCCTATATGGGGAACAATGACTTAACCATGTGGCTAGGGGAAAAGTTGG 1462
           |  |            ||||||      || ||||||  |    |||   |  |  | |
Db     528 ACACACAACGTTTAATTATGGGATTTCGAGATTTAACCCAGGTTTTAGCACATGATCTAG 587

Qy    1463 GTTTCTATACAGACTTTGGTCCAAGCACAAGGTATCACACTTGGGGGATTATGGCTTTGT 1522
           |  | ||| | || | ||| ||| || |     | || ||||||||   |  |   | |
Db     588 GAATGTATGCGGATTATGGCCCAGGCCCTGATAAACATACTTGGGGCGCTGCGTTACTCT 647

Qy    1523 CAAGATACCCAATTGTGAAATCTGAGCATCACCTTCTTCCGTCACCAGAGGGCGAGATCG 1582
           | | ||  || ||||| || ||    || ||  |  | || || ||  | || ||  | |
Db     648 CCAAATTTCCTATTGTCAACTCGACTCACCATTTGTTACCTTCTCCTCAAGGAGAACTTG 707

Qy    1583 CACCAGCCATCACATTGACCGTTAACATTTCGGGCAAGCTGGTGGATTTTGTCGTGACAC 1642
           | || || ||       ||  |  |  | |  ||  |||| | ||| |||| ||  | |
Db     708 CTCCTGCTATTCATGCAACACTAGATGTCTACGGAGAGCTAATAGATGTTGTGGTTTCTC 767

Qy    1643 ACTTTGGGAACCACGAAGATGACCTCGACAGGAAACTGCAGGCTATTGCTGTTTCAAAAC 1702
           ||  |||  |  | ||| | |  | ||  |    || |||   | ||  |  |   |
Db     768 ACAATGGTCAATATGAAAGTCAGTTGGATCGACGTCTTCAGAGCACTGAGTTGGCTCGAA 827

Qy    1703 TACTGAAAAGTAGCTCTAATCAAGTGATATTTCTGGGATATATCACTTCAGCACCTGGCT 1762
           |  ||      ||  | |  |   |  | ||||| |||||| |   ||||     ||||
Db     828 TTATGCGCGAGAGTCCAAGGCCTCTAGTGTTTCTCGGATATGTTGTTTCAAACGTTGGCC 887

Qy    1763 CCAGAGATTATCTACAGCTCACTGAACATGGCAATGTGAAGGATATCGACAGCACTGATC 1822
           |    |       || ||   | |        ||   || || ||     ||||
Db     888 AAGAACCCCAAACTATTCTAACGCGAGACACGGGAATGTTAGACATTGAACCTGCTGACT 947

Qy    1823 ATGACAGATGGTGTGAATACATTATGTATCGAGGGCTGATCAGGTTGGGTTATGCAAGAA 1882
           ||||  | |||||| | || ||  | |||||||||  | |   |  | ||||||| |||
Db     948 ATGATCGTTGGTGTCAGTATATATTTTATCGAGGTGTCAAGCGAATTGGTTATGCTAGAC 1007

Qy    1883 TCTCCCATGCTGAACTGAGTGATTCAGAAATTCAGATGGCAAAATTTAGGATCCCTGATG 1942
           |   |   ||   | |  ||  | ||  |||| |  | ||| |||  | |  |  | |
Db    1008 TTCATCGCTCTACTATAACAGACACCGAGCTTCAAACCGGAAAGTTTTTGGTTACCAAAG 1067

Qy    1943 A 1943
           |
Db    1068 A 1068


RESULT 8
SPAC688
LOCUS       SPAC688                34034 bp    DNA     linear   PLN 08-MAY-2000
DEFINITION  S.pombe chromosome I cosmid c688.
ACCESSION   AL355632
VERSION     AL355632.1  GI:7768476
KEYWORDS    Actin-related protein; cell polarization; coiled-coil; cut9
            interacting protein; dna polymerase zeta catalytic subunit; drug
            sensitivity; DUF51; endocytosis internalization phase; ENTH domain;
            glutathione S transferase; I/LWEQ domain; Kornberg's mediator (SRB)
            subcomplex; membrane cytoskeleton assembly; microtubule stability;
            mitochondrial carrier protein; Protein of unknown function domain;
            RNA polymerase II holoenzyme; scn1; Tf1-type LTR; yeast SLA2; yeast
            SRB8; yeast YCR017C.
SOURCE      fission yeast.
  ORGANISM  Schizosaccharomyces pombe
            Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
            Schizosaccharomycetales; Schizosaccharomycetaceae;
            Schizosaccharomyces.
REFERENCE   1  (bases 1 to 34034)
  AUTHORS   Aert,R., Robben,J., Volckaert,G., Wood,V., Rajandream,M.A. and
            Barrell,B.G.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-APR-2000) European Schizosaccharomyces genome
            sequencing project, Sanger Centre, The Wellcome Trust Genome
            Campus, Hinxton, Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk
            and Katholieke Universiteit Leuven, Laboratory of Gene Technology,
            Kardinaal Mercierlaan 92, B-3001 Leuven, Belgium
COMMENT     Notes:
            Details of yeast sequencing at the Sanger Centre are available on
            the World Wide Web.
            (URL, http://www.sanger.ac.uk/Projects/S_pombe/)
            During 1995 to 1996 about 66% of S. pombe chromosome 1 was
            sequenced by the Sanger Centre.  The sequencing of the S. pombe
            genome is now being continued with funding from The European
            Commission. Fourteen European sequencing laboratories, including
            the Sanger Centre, are participating in the project.
            Protein coding regions (CDS) have been predicted with the help of
            computer analysis using the Genefinder program in PomBase (an ACEDB
            database) with additional predictions for the branch-acceptor sites
            supplied by the program Sp3splice. CAUTION: It is possible that for
            any individual CDS we may have underestimated or overestimated the
            number of introns/exons or we may not have chosen the correct
            splice donor/acceptor sites.
            CDS are numbered using the following system eg SPBC25H2.01c. SP (S.
            pombe), B (chromosome 2), c25H2 (cosmid name), .01 (first CDS), c
            (complementary strand).
            The more significant matches with motifs in the PROSITE database
            are also included but some of these may be fortuitous.
            The length in codons is given for each CDS.
            IMPORTANT: This sequence MAY NOT be the entire insert of the
```

```
        sequenced clone.  It may be shorter because we only sequence
        overlapping sections once, or longer, because we arrange for a
        small overlap between neighbouring submissions.
        Cosmid c688 is overlapped at the 3' end by cosmid c3G9, EMBL entry
        SPAC3G9, accession number AL021046.
FEATURES             Location/Qualifiers
     source          1. .34034
                     /organism="Schizosaccharomyces pombe"
                     /strain="972h-"
                     /db_xref="taxon:4896"
                     /chromosome="I"
                     /map="IL"
                     /clone="cosmid c688"
     gene            1. .1280
                     /gene="SPAC688.01"
     CDS             1. .1280
                     /partial
                     /gene="SPAC688.01"
                     /note="SPAC688.01, len:425, SIMILARITY:Saccharomyces
                     cerevisiae, YCQ7_YEAST, hypothetical 107.9 kd protein.,
                     (953 aa), fasta scores: opt: 1034, E():0, (46.5% identity
                     in 430 aa)"
                     /codon_start=3
                     /label=SPAC688.01
                     /product="hypothetical protein similar to yeast YCR017C
                     protein involved in sensitivity to certain drugs"
                     /protein_id="CAB90768.1"
                     /db_xref="GI:7768477"
                     /translation="SYFMYINHGWCSYLGGLIFTSYVLIYSFASIRISSFYSPAKVWG
                     GAFLVYILYSLAHVWVVAYEFVPGGPILRERTSYILIFIGWNLAALVPAYSGESKEPN
                     KADSSVVDIKQSDSSYRRRSFKKSLLTGFCLALMALKFAIQNMPPYDYTPYHPNEKLF
                     TAGIWTIHFGLDNFMYASENRIRDAVRDMELDVFGLLESDTQRLIMGFRDLTQVLAHD
                     LGMYADYGPGPDKHTWGAALLSKFPIVNSTHHLLPSPQGELAPAIHATLDVYGELIDV
                     VVSHNGQYESQLDRRLQSTELARIMRESPRPLVFLGYVVSNVGQEPQTILTRDTGMLD
                     IEPADYDRWCQYIFYRGVKRIGYARLHRSTITDTELQTGKFLVTKDLGRNVRIDKEHV
                     PESHRYPSLFEGTGVNGHYYDNNLVVHEPWYYD"
     gene            complement(join(1545. .1979,2109. .2256,2343. .2392))
                     /gene="SPAC688.02c"
     CDS             complement(join(1545. .1979,2109. .2256,2343. .2392))
                     /gene="SPAC688.02c"
                     /note="SPAC688.02c, len:210, SIMILARITY:Saccharomyces
                     cerevisiae, Q12143, chromosome xvi reading frame orf
                     ypl233w., (216 aa), fasta scores: opt: 183, E():9.1e-05,
                     (22.7% identity in 185 aa)"
                     /codon_start=1
                     /label=SPAC688.02c
                     /product="hypothetical protein with very low similarity to
                     yeast ypl233w"
                     /protein_id="CAB90769.1"
                     /db_xref="GI:7768478"
                     /translation="MSSHGDASVLPKIALESINDYSYVKSVFRQAIQEKLAIHLPEQA
                     SREAGMGREDQLRIRVKEMLNELIESTFSIIQENVTINGFDANSALQDPKNSDQIEPF
                     DLALRTRVQQLFNEVEDAHVLVARYRKSVPAQYEKAYVDAMEQQTAFLRNVKDDYVSL
                     QDKEVENPDEQTSTTVFRKEDIDRYEQTIAKVAYLKKNLPRVVARLEKTP"
     misc_feature    complement(1980. .1990)
                     /gene="SPAC688.02c"
                     /note="ttaacatatag, splice branch and acceptor"
     misc_feature    complement(2103. .2108)
                     /gene="SPAC688.02c"
                     /note="gtgagt, splice donor sequence"
     misc_feature    complement(2257. .2271)
                     /gene="SPAC688.02c"
                     /note="ttaacactttataag, splice branch and acceptor"
     misc_feature    complement(2337. .2342)
                     /gene="SPAC688.02c"
                     /note="gtgtgt, splice donor sequence"
     gene            complement(join(2596. .2953,3083. .3297,3381. .3422))
                     /gene="SPAC688.03c"
     CDS             complement(join(2596. .2953,3083. .3297,3381. .3422))
                     /gene="SPAC688.03c"
                     /note="SPAC688.03c, len:204, SIMILARITY:Homo sapiens,
                     Q9Y4X0, ammecr1 protein., (333 aa), fasta scores: opt:
                     443, E():3.4e-25, (40.3% identity in 186 aa)"
                     /codon_start=1
                     /label=SPAC688.03c
                     /product="conserved hypothetical protein"
                     /protein_id="CAB90770.1"
                     /db_xref="GI:7768479"
                     /translation="MICWPSLNKESWNDKEYCYYCFEVVAATLEHRKVRDKWNAKSWT
                     RSIPLFVKFASGKGHDKQLRGCIGTFRARPLVTNLTYFSKQAAFCDERFRPISLGELA
                     LLECQIDLLVDFEPIDDPDLDWEVGIHGVSIKFTANGIRYSSTYLPSVAAEQRWDQEET
                     LESLIHKAGYYGSIRSLQITATRYKSLEIGCTYEEYLHNLELLG"
     misc_feature    complement(join(2662. .2953,3083. .3273))
                     /gene="SPAC688.03c"
                     /note="Match to PF01871 DUF51, Protein of unknown function
                     Score 49.41"
     misc_feature    complement(2954. .2976)
                     /gene="SPAC688.03c"
                     /note="ctaatacaacattttaaaaatag, splice branch and
                     acceptor"
     misc_feature    complement(3077. .3082)
                     /gene="SPAC688.03c"
                     /note="gtaagt, splice donor sequence"
     misc_feature    complement(3298. .3315)
                     /gene="SPAC688.03c"
                     /note="ctaatataaatggacaag, splice branch and acceptor"
     misc_feature    complement(3375. .3380)
                     /gene="SPAC688.03c"
                     /note="gtatga, splice donor sequence"
     gene            complement(4047. .4775)
                     /gene="SPAC688.04c"
     CDS             complement(4047. .4775)
                     /gene="SPAC688.04c"
                     /note="SPAC688.04c, len:242, SIMILARITY:Saccharomyces
                     cerevisiae, YIV8_YEAST, Glutathione transferase GTT1, (234
                     aa), fasta scores: opt: 274, E():3.5e-11, (27.0% identity
                     in 222 aa)"
                     /codon_start=1
                     /label=SPAC688.04c
                     /product="putative glutathione S transferase"
                     /protein_id="CAB90771.1"
                     /db_xref="GI:7768480"
                     /translation="MIVLHHLKNSRSTRIVWMLEELKVPYEIKVYDRVDQRAPPAYTK
                     LSPLGKSPIVVDDGVTYIESAAILEHLVRKYGPSFKPSEEDVAELEKYELWMHFSEAS
                     LMPFIWASHVLDLSVNMTPIFFRYIVRQFVNGIKSKYLSKETFLNLDYIDNHLASNEY
                     FAGEQFTAADPQMCFPIFAAQRDYLSQKPYKNIKRWMRVVSDRPACRIAAEKVEDNTL
                     TLFSDVERYSHPPTPPPEQVRSDE"
     misc_feature    complement(4164. .4775)
                     /gene="SPAC688.04c"
                     /note="Match to PF00043 GST, Glutathione S-transferases.
                     Score 85.13"
     gene            complement(join(5637. .6842,6927. .6980))
                     /gene="SPAC688.06c"
     CDS             complement(join(5637. .6842,6927. .6980))
                     /gene="SPAC688.06c"
                     /note="SPAC688.06c, len:419, predicted to contain
                     coiled-coil region, SIMILARITY:S. cerevisiae, LOW to
                     Actin-related protein, regulator of microtubule stability,
                     YB61_YEAST, (324 aa), fasta scores: opt: 125, E():0.76,
                     (21.3% identity in 300 aa)"
                     /codon_start=1
                     /label=SPAC688.06c
                     /product="hypothetical protein very putative Actin-related
                     protein, regulator of microtubule stability"
                     /protein_id="CAB90772.1"
                     /db_xref="GI:7768481"
                     /translation="MSAEEYIEVSSSPDIFTDDDDMITIEPELNKNPKDCNSKRKRSV
                     TECCEIRLITSKCDFESTQQLVHHNCTGHKVHEHNLNAVDEEDFDTENEPLLFSSFSD
                     NESDILEPDLNTRVAEDNDVLLSRYSKIKNSASCRNTFEHSAYHSNREEISSSGFYYH
                     RKPQLFEKSLEKLGNKSIEANRSPLIKELCESANSTENVCFSVSTVDEIQQRHPSAGH
                     SIDSTCQSNSFLEGDSATHKKKKTDNIKEFTSCEFNDRSRTLLNYAGYMDTNKNADNE
                     AKSLKEKLENFPVEKLRAIAESYGFKSSDSKATLIKIVESCLDAIDSRSQSKKLGKET
                     PHDYLITSTKTVLEFDDIVTQTHRAISQVVKQAKDNSVWIKILTYSAIDVEEFQLWLK
                     RKNLNVSLDLIKSWCDKYGVLMKGSWH"
     misc_feature    complement(6843. .6857)
                     /gene="SPAC688.06c"
```

```
                /note="ttaactgagctttag, splice branch and acceptor"
     misc_feature    complement(6921. .6926)
                /gene="SPAC688.06c"
                /note="gtatga, splice donor sequence"
     gene            complement(7359. .10475)

  Query Match             5.8%;  Score 121.8;  DB 8;  Length 34034;
  Best Local Similarity   49.0%;  Pred. No. 7e-23;
  Matches  324;  Conservative    0;  Mismatches  337;  Indels    0;  Gaps    0;

Qy    1283 TCTCTGCTGCCATCTGGCCTTTCAGGTTTGGATATGACAATGAAGGGTGGTCTAGTCTAG 1342
           || |||||| || ||| || |    ||||| |  ||||||     ||   | |||
Db     481 TCACTGCTGGTATTTGGACTATTCACTTTGGCTTAGACAATTTCATGTATGCAAGTGAGA 540

Qy    1343 AAAGATCAGCTCACCTGCTCAATGAAACAGGTGCAGATTTCATAACAATTTTGGAGAGTG 1402
           |  |   |    |      |    || |  |     ||| || |     |  ||||   ||
Db     541 ATCGTATACGCGATGCTGTTCGAGATATGGAACTTGATGTCTTTGGTCTACTGGAATCTG 600

Qy    1403 ATGCTTCTAAGCCCTATATGGGGAACAATGACTTAACCATGTGGCTAGGGGAAAAGTTGG 1462
           |  |           ||||||       || ||||||  |    |||   |  |  | |
Db     601 ACACACAACGTTTAATTATGGGATTTCGAGATTTAACCCAGGTTTTAGCACATGATCTAG 660

Qy    1463 GTTTCTATACAGACTTTGGTCCAAGCACAAGGTATCACACTTGGGGGATTATGGCTTTGT 1522
           |  | ||| | || | ||| ||| || |     | || ||||||||  |  |    | |
Db     661 GAATGTATGCGGATTATGGCCCAGGCCCTGATAAACATACTTGGGGCGCTGCGTTACTCT 720

Qy    1523 CAAGATACCCAATTGTGAAATCTGAGCATCACCTTCTTCCGTCACCAGAGGGCGAGATCG 1582
           | | ||  || ||||| || ||    || ||  |  | || || ||  | || ||  | |
Db     721 CCAAATTTCCTATTGTCAACTCGACTCACCATTTGTTACCTTCTCCTCAAGGAGAACTTG 780

Qy    1583 CACCAGCCATCACATTGACCGTTAACATTTCGGGCAAGCTGGTGGATTTTGTCGTGACAC 1642
           | || || ||       ||  |  |  | |  ||  |||| | ||| |||| ||  | |
Db     781 CTCCTGCTATTCATGCAACACTAGATGTCTACGGAGAGCTAATAGATGTTGTGGTTTCTC 840

Qy    1643 ACTTTGGGAACCACGAAGATGACCTCGACAGGAAACTGCAGGCTATTGCTGTTTCAAAAC 1702
           ||  |||  |  | |||  | |  | ||  |    || |||   | ||   |  |   |
Db     841 ACAATGGTCAATATGAAAGTCAGTTGGATCGACGTCTTCAGAGCACTGAGTTGGCTCGAA 900

Qy    1703 TACTGAAAAGTAGCTCTAATCAAGTGATATTTCTGGGATATATCACTTCAGCACCTGGCT 1762
           |  ||      ||  | |  |   |  | ||||| |||||| |   ||||     ||||
Db     901 TTATGCGCGAGAGTCCAAGGCCTCTAGTGTTTCTCGGATATGTTGTTTCAAACGTTGGCC 960

Qy    1763 CCAGAGATTATCTACAGCTCACTGAACATGGCAATGTGAAGGATATCGACAGCACTGATC 1822
           |   |       || ||   | |         ||   || || ||     ||||
Db     961 AAGAACCCCAAACTATTCTAACGCGAGACACGGGAATGTTAGACATTGAACCTGCTGACT 1020

Qy    1823 ATGACAGATGGTGTGAATACATTATGTATCGAGGGCTGATCAGGTTGGGTTATGCAAGAA 1882
           ||||  | |||||| | || ||  | ||||||||  | |   |  | ||||||| |||
Db    1021 ATGATCGTTGGTGTCAGTATATATTTTATCGAGGTGTCAAGCGAATTGGTTATGCTAGAC 1080

Qy    1883 TCTCCCATGCTGAACTGAGTGATTCAGAAATTCAGATGGCAAAATTTAGGATCCCTGATG 1942
           |    |   ||    | |  ||  | ||  |||| |  | ||| |||  | |  |  | |
Db    1081 TTCATCGCTCTACTATAACAGACACCGAGCTTCAAACCGGAAAGTTTTTGGTTACCAAAG 1140

Qy    1943 A 1943
           |
Db    1141 A 1141


RESULT 9
AX337478/c
LOCUS       AX337478                305 bp    DNA     linear   PAT 09-JAN-2002
DEFINITION  Sequence 7987 from Patent WO0194629.
ACCESSION   AX337478
VERSION     AX337478.1  GI:18128197
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Young,P.E., Augustus,M., Carter,K.C., Ebner,R., Endress,G.,
            Horrigan,S., Soppet,D.R. and Weaver,Z.
  TITLE     Cancer gene determination and therapeutic screening using signature
            gene sets
  JOURNAL   Patent: WO 0194629-A 7987 13-DEC-2001;
            Avalon Pharmaceuticals (US)
FEATURES             Location/Qualifiers
     source          1. .305
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT       81 a     45 c     44 g    127 t      8 others
ORIGIN

  Query Match             5.3%;  Score 112;  DB 6;  Length 305;
  Best Local Similarity   94.1%;  Pred. No. 2.5e-20;
  Matches  112;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy    1985 ACAGAGAAGTTTCTGAGAAAATTCATTTTAATCCCAGATTTGGATCCTACAAAGAAGGAC 2044
           || |||||||||||| |||||||||||||||| |||||||||||||||||||||| |
Db     305 ACNGAGAAGTTTCTGNGAAAATTCATTTTAATCCCNGATTTGGATCCTACAAAGAAGGNC 246

Qy    2045 ACAATTATGAAAACAACCATCATTTTCATATGAATACTCCCAAATACTTTTTATGAAAC 2103
           |||||||||||||||||||||||| |||||||||||||||||||||| || ||||||
Db     245 ACAATTATGAAAACAACCATCATTNTCATATGAATACTCCCAAATACTTNTTNTGAAAC 187


RESULT 10
CNS06JWV
LOCUS       CNS06JWV                995 bp    DNA     linear   STS 10-JAN-2001
DEFINITION  T3 end of clone XAS0AA002B08 of library XAS0AA from strain CLIB 533
            of Saccharomyces bayanus, sequence tagged site.
ACCESSION   AL402149
VERSION     AL402149.1  GI:12160780
KEYWORDS    STS.
SOURCE      Saccharomyces bayanus.
  ORGANISM  Saccharomyces bayanus
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE   1  (bases 1 to 995)
  AUTHORS   Bon,E., Neuveglise,C., Casaregola,S., Artiguenave,F., Wincker,P.,
            Aigle,M. and Durrens,P.
  TITLE     Genomic Exploration of the Hemiascomycetous Yeasts: 5.
            Saccharomyces bayanus var. uvarum
  JOURNAL   FEBS Lett. 487 (1), 37-41 (2000)
   PUBMED   11152880
REFERENCE   2  (bases 1 to 995)
  AUTHORS   Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
            Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,
            de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B.,
            Malpertuy,A., Neuveglise,C., Ozier-Kalogeropoulos,O., Potier,S.,
            Saurin,W., Tekaia,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,
            Wincker,P. and Weissenbach,J.
  TITLE     Genomic Exploration of the Hemiascomycetous Yeasts: 1. A set of
            yeast species for molecular evolution studies(1)
  JOURNAL   FEBS Lett. 487 (1), 3-12 (2000)
   PUBMED   11152876
REFERENCE   3  (bases 1 to 995)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-SEP-2000) Genoscope - Centre National de Sequencage,
            2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
            seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT     This STS is part of a random genomic sequencing program of thirteen
            yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
            exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
            Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
            lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
            angusta, Debaryomyces hansenii var.  hansenii, Pichia sorbitophila,
            Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
            5 kb were prepared and both extremities were sequenced. See
            keywords for description of this sequence and for the sequence of
            the other extremity of this insert.
FEATURES             Location/Qualifiers
     source          1. .995
                     /organism="Saccharomyces bayanus"
```

```
                     /strain="CLIB 533"
                     /variety="uvarum"
                     /db_xref="taxon:4931"
                     /clone="XAS0AA002B08"
                     /clone_lib="XAS0AA"
                     /note="end : T3"
     misc_feature    <3. .>958
                     /note="similar to Saccharomyces cerevisiae ORF YCR017c [
                     similarity to hypothetical S.pombe protein ]
                     1 putative frameshift(s)"
                     /evidence=not_experimental
BASE COUNT      276 a    190 c    210 g    318 t      1 others
ORIGIN

  Query Match             5.3%;  Score 110.6;  DB 11;  Length 995;
  Best Local Similarity   52.2%;  Pred. No. 7e-20;
  Matches  314;  Conservative    1;  Mismatches  280;  Indels    6;  Gaps    3;

Qy   1277 AAGAGGTCTCTGCTGCCATCTGGCCTTTCAGGTTTGGATATGACAATGAAGGGTGGTCTA 1336
           ||  | |||||||||  || ||| |  ||   |||||   ||| |||||   |||| |
Db    391 AATTGATCTCTGCTGGTATTTGGACCATCCATTTTGGTCTTGATAATGATATGTGGGCAT 450

Qy   1337 GTCTAGAAAGATCAGCTCACCTGCTCAATGAAACAGGTGCAGATTTCATAACAATTTTGG 1396
           |  ||| |||       || |  | || ||||  |     ||| |  ||  | |  | |
Db    451 CTGAAGACAGAATGATCAACTTAATAAAAGAAATGGAGTTGGATGTTGTAGGATTACTAG 510

Qy   1397 AGAGTGATGCTTCTAAGCCCTATATGGGGAACAATGACTTAACCATGTGGCTAGGGGAAA 1456
           | |  ||| |    |       ||||||  |||  ||  ||||||     ||||   |
Db    511 AAACAGATACGCAAAGAATTATTATGGGTAACAGGGATCTAACCAGTAAACTAGCCCATG 570

Qy   1457 AGTTGGGTTTCTATACAGACTTTGGTCCAAGCACAAGGTATCACACTTGGGGGATTATGG 1516
           |  ||  | | ||| | || |||||  ||   |  |||   | |||||  |||||   |||
Db    571 ATCTGAATATGTATGCTGATTTTGGCCCGGGTCCAAACAAACACACATGGGGCTGTATTC 630

Qy   1517 CTTTGTCAAGATACCCAATTGTGAAATCTGAGCATCACCTTCTTCCGTCACCAGAGGGCG 1576
           | | ||  | ||  ||| ||  | ||  |   ||| || ||  | || || || | ||| |
Db    631 TTCTTTCCAAATTCCCTATCATTAATGCAACGCACCATCTATTGCCATCCCCTGTGGGAG 690

Qy   1577 AGATCGCACCAGCCATCACATTGACCGTTAACA---TTTCGGGCAAGCTGGTGGATTTTG 1633
           || | || || |||||       ||  || | |    |   | || | | || ||| |
Db    691 AGCTTGCGCCGGCCATTCACGCCACTCTTCAAACGTATAATGACACGTTAGTTGATGTCT 750

Qy   1634 TCGTGACACACTTTGGGAACCACGAAGATGACCTCGACAGGAAACTGCAGGCTATTGCTG 1693
           | ||    ||  |||  |  | ||||||||    || || | | | ||   || |
Db    751 TTGTCTTCCATAGTGGACAAGAGGAAGATGAAGAAGATAGAAGATTACAAAGTAATTACA 810

Qy   1694 TTTCAAAACTACTGAAAAGTAGCTCTAATCAAGTGATATTTCTGGGATATATCACTTCAG 1753
           |  | |||  |  ||  || |    |    | ||  || ||  |  | ||  |       |
Db    811 TGGCTAAATTGATGGGAAATTCGACCCGACCAGCTATTTTATTAAGTTACTTGGTGGTTG 870

Qy   1754 CACCTGGCTCCAGAGATTATCTACAGCTCACT--GAACATGGCAATGTGAAGGATATCGA 1811
           || |||    |  | ||      ||    |  | |       || | || || ||
Db    871 ACCCGGGCGAAGGTAACTACAATACGCATGTTAGTGAACTTCTGGCATGCATGACATTGA 930

Qy   1812 CAGCACTGATCATGACAGATGGTGTGAATACATTATGTATCGAGGGCTGATCAGGTTGGG 1871
           |  |||||  |||| || ||||||| ||| ||  |||||  ||   |||  ||   ||:
Db    931 CCCTACTGACGATGATAGGTGGTGTG-ATATATCTTGTATAAAGATTTGAGAAGACAGGK 989

Qy   1872 T 1872
           |
Db    990 T 990


RESULT 11
AC106576/c
LOCUS       AC106576              165893 bp    DNA     linear   HTG 12-JAN-2002
DEFINITION  Rattus norvegicus clone CH230-198N10, *** SEQUENCING IN PROGRESS
            ***, 68 unordered pieces.
ACCESSION   AC106576
VERSION     AC106576.1  GI:18139100
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 165893)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J.,
            Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
            Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
            Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
            Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
            Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
            Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
            Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
            Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
            Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
            Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
            Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
            Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
            Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
            Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
            Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
            Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
            Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
            Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
            Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
            Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
            Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
            Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
            Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
            Morgan,M., Morris,S., Moser,M., Neal,D., Newtson,J., Newtson,N.,
            Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
            Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
            Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
            Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M.,
            Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N.,
            Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H.,
            Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
            Tang,H., Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N.,
            Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R.,
            Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
            Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S.,
            Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 165893)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-JAN-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     -------------- Genome Center
               Center: Baylor College of Medicine
               Center code: BCM
               Web site: http://www.hgsc.bcm.tmc.edu/
               Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
               Center project name: GLJY
               Center clone name: CH230-198N10
            -------------- Summary Statistics
               Assembly program: Phrap; version 0.990329First call to
            findPhrapList
               Consensus quality: 137306 bases at least Q40
               Consensus quality: 144043 bases at least Q30
               Consensus quality: 149677 bases at least Q20
               Estimated insert size: 139846; sum-of-contigs estimation
               Quality coverage: 0x in Q20 bases; agarose-fp estimation
               Quality coverage: 2.2x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 68 contigs. The true order of the pieces
```

```
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*       1     7956: contig of 7956 bp in length
*    7957     8056: gap of unknown length
*    8057    13903: contig of 5847 bp in length
*   13904    14003: gap of unknown length
*   14004    17958: contig of 3955 bp in length
*   17959    18058: gap of unknown length
*   18059    22354: contig of 4296 bp in length
*   22355    22454: gap of unknown length
*   22455    26589: contig of 4135 bp in length
*   26590    26689: gap of unknown length
*   26690    31248: contig of 4559 bp in length
*   31249    31348: gap of unknown length
*   31349    36556: contig of 5208 bp in length
*   36557    36656: gap of unknown length
*   36657    40383: contig of 3727 bp in length
*   40384    40483: gap of unknown length
*   40484    44415: contig of 3932 bp in length
*   44416    44515: gap of unknown length
*   44516    49038: contig of 4523 bp in length
*   49039    49138: gap of unknown length
*   49139    52349: contig of 3211 bp in length
*   52350    52449: gap of unknown length
*   52450    56176: contig of 3727 bp in length
*   56177    56276: gap of unknown length
*   56277    60332: contig of 4056 bp in length
*   60333    60432: gap of unknown length
*   60433    62803: contig of 2371 bp in length
*   62804    62903: gap of unknown length
*   62904    66621: contig of 3718 bp in length
*   66622    66721: gap of unknown length
*   66722    69020: contig of 2299 bp in length
*   69021    69120: gap of unknown length
*   69121    72220: contig of 3100 bp in length
*   72221    72320: gap of unknown length
*   72321    75607: contig of 3287 bp in length
*   75608    75707: gap of unknown length
*   75708    78960: contig of 3253 bp in length
*   78961    79060: gap of unknown length
*   79061    81780: contig of 2720 bp in length
*   81781    81880: gap of unknown length
*   81881    85140: contig of 3260 bp in length
*   85141    85240: gap of unknown length
*   85241    87883: contig of 2643 bp in length
*   87884    87983: gap of unknown length
*   87984    90762: contig of 2779 bp in length
*   90763    90862: gap of unknown length
*   90863    93086: contig of 2224 bp in length
*   93087    93186: gap of unknown length
*   93187    96030: contig of 2844 bp in length
*   96031    96130: gap of unknown length
*   96131    98070: contig of 1940 bp in length
*   98071    98170: gap of unknown length
*   98171   100910: contig of 2740 bp in length
*  100911   101010: gap of unknown length
*  101011   102968: contig of 1958 bp in length
*  102969   103068: gap of unknown length
*  103069   104748: contig of 1680 bp in length
*  104749   104848: gap of unknown length
*  104849   106306: contig of 1458 bp in length
*  106307   106406: gap of unknown length
*  106407   108402: contig of 1996 bp in length
*  108403   108502: gap of unknown length
*  108503   110234: contig of 1732 bp in length
*  110235   110334: gap of unknown length
*  110335   112075: contig of 1741 bp in length
*  112076   112175: gap of unknown length
*  112176   113608: contig of 1433 bp in length
*  113609   113708: gap of unknown length
*  113709   115299: contig of 1591 bp in length
*  115300   115399: gap of unknown length
*  115400   116670: contig of 1271 bp in length
*  116671   116770: gap of unknown length
*  116771   119076: contig of 2306 bp in length
*  119077   119176: gap of unknown length
*  119177   120752: contig of 1576 bp in length
*  120753   120852: gap of unknown length
*  120853   122336: contig of 1484 bp in length
*  122337   122436: gap of unknown length
*  122437   125169: contig of 2733 bp in length
*  125170   125269: gap of unknown length
*  125270   126815: contig of 1546 bp in length
*  126816   126915: gap of unknown length
*  126916   128187: contig of 1272 bp in length
*  128188   128287: gap of unknown length
*  128288   129839: contig of 1552 bp in length
*  129840   129939: gap of unknown length
*  129940   131660: contig of 1721 bp in length
*  131661   131760: gap of unknown length
*  131761   133894: contig of 2134 bp in length
*  133895   133994: gap of unknown length
*  133995   135658: contig of 1664 bp in length
*  135659   135758: gap of unknown length
*  135759   136862: contig of 1104 bp in length
*  136863   136962: gap of unknown length
*  136963   138366: contig of 1404 bp in length
*  138367   138466: gap of unknown length
*  138467   139604: contig of 1138 bp in length
*  139605   139704: gap of unknown length
*  139705   141083: contig of 1379 bp in length
*  141084   141183: gap of unknown length
*  141184   142432: contig of 1249 bp in length
*  142433   142532: gap of unknown length
*  142533   144347: contig of 1815 bp in length
*  144348   144447: gap of unknown length
*  144448   145945: contig of 1498 bp in length
*  145946   146045: gap of unknown length
*  146046   147800: contig of 1755 bp in length
*  147801   147900: gap of unknown length
*  147901   149217: contig of 1317 bp in length
*  149218   149317: gap of unknown length
*  149318   150337: contig of 1020 bp in length
*  150338   150437: gap of unknown length
*  150438   151919: contig of 1482 bp in length

  Query Match             3.9%;  Score 81.6;  DB 2;  Length 165893;
  Best Local Similarity   78.2%;  Pred. No. 1.7e-11;
  Matches  111;  Conservative    0;  Mismatches   29;  Indels    2;  Gaps    1;

Qy    1882 ATCTCCCATGCTGAACTGAGTGATTCAGAAATTCAGATGGCAAAATTTAGGATCCCTGAT 1941
           || || ||||||||||||||||| || || ||       || ||||| |||||||| |||
Db  165892 ATTTCTCATGCTGAACTGAGTGACTCTGAGAT--CAGATGCCAAATTCAGGATCCCAGAT 165835

Qy    1942 GACCCCACTAATTATAGAGACAACCAGAAAGTGGTCATAGACCACAGAGAAGTTTCTGAG 2001
           || |||   ||||| |||||||||||||| ||| ||||||||||||||| |||| |  ||
Db  165834 GATCCCTTCAATTACAGAGACAACCAGAGAGTCGTCATAGACCACAGAGGAGTTCCCAAG 165775

Qy    2002 AAAATTCATTTTAATCCCAGAT 2023
           || |||||||| || || || |
Db  165774 AACATTCATTTCAACCCGAGGT 165753


RESULT 12
HUMTPIPSA/c
LOCUS       HUMTPIPSA               1348 bp    DNA     linear   PRI 14-JAN-1995
DEFINITION  Human TPI (triosephosphate isomerase) pseudogene psi-13C.
ACCESSION   K03223
VERSION     K03223.1  GI:339844
KEYWORDS    pseudogene; triose-phosphate isomerase.
SOURCE      Human DNA, clone psi-hTPI-13C.
  ORGANISM  Homo sapiens
```

```
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1348)
  AUTHORS   Brown,J.R., Daar,I.O., Krug,J.R. and Maquat,L.E.
  TITLE     Characterization of the functional gene and several processed
            pseudogenes in the human triosephosphate isomerase gene family
  JOURNAL   Mol. Cell. Biol. 5 (7), 1694-1706 (1985)
  MEDLINE   85267686
COMMENT     The intronless pseudogene presented below, is missing 317 bases
            between positions 725 and 726, eliminating the stop codon.  The
            start codon is located at positions 264-266.  This gene is flanked
            by short direct repeats, 'taaattt', suggesting that it is generated
            by integration of a reverse-transcribed copy of TPI mRNA into germ
            line DNA via a transposon-like mechanism.  A potential poly-A
            signal is located at positions 1119-1124.
FEATURES             Location/Qualifiers
     source          1. .1348
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /map="12p13"
     misc_feature    1
                     /gene="TPI1"
                     /note="G00-119-617"
BASE COUNT      385 a    313 c    304 g    346 t
ORIGIN      262 bp upstream of NcoI site.

  Query Match             3.9%;  Score 81.4;  DB 9;  Length 1348;
  Best Local Similarity   98.8%;  Pred. No. 1.2e-11;
  Matches   82;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy    1187 TTCTGTGGCTGCTTGTTGGTGTGGGATTGTTGGGATTAGGACTACGGCATAAAGCCTATG 1246
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     169 TTCTGTGGCTGCTTGTTGGTGTGGGATTGTTGGGATTAGGACTACGGCATAAAGCCTATG 110

Qy    1247 AGAGAAAACTGGGCAAAGTGGCA 1269
           |||||||||||||||||||||| |
Db     109 AGAGAAAACTGGGCAAAGTGGTA 87


RESULT 13
AX106346/c
LOCUS       AX106346                 54 bp    DNA     linear   PAT 30-APR-2001
DEFINITION  Sequence 127 from Patent WO0125272.
ACCESSION   AX106346
VERSION     AX106346.1  GI:13922028
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 54)
  AUTHORS   Xu,J., Skeiky,Y.A., Reed,S.G. and Cheever,M.A.
  TITLE     Compositions and methods for therapy and diagnosis of prostate
            cancer
  JOURNAL   Patent: WO 0125272-A 127 12-APR-2001;
            CORIXA CORPORATION (US)
FEATURES             Location/Qualifiers
     source          1. .54
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT       23 a     17 c      9 g      5 t
ORIGIN

  Query Match             2.6%;  Score 54;  DB 6;  Length 54;
  Best Local Similarity   100.0%;  Pred. No. 0.00047;
  Matches   54;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     724 CTGCTGTGCTTGGCAAGTGGATTGATGCTTCCATCTTGTTTGTGGTTTCGTGGT 777
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      54 CTGCTGTGCTTGGCAAGTGGATTGATGCTTCCATCTTGTTTGTGGTTTCGTGGT 1


RESULT 14
AX140637/c
LOCUS       AX140637                 54 bp    DNA     linear   PAT 31-MAY-2001
DEFINITION  Sequence 127 from Patent WO0134802.
ACCESSION   AX140637
VERSION     AX140637.1  GI:14280751
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 54)
  AUTHORS   Xu,J., Dillon,D.C., Mitcham,J.L., Harlocker,S.L., Jiang,Y.,
            Reed,S.G., Kalos,M.D., Retter,M.W., Stolk,J.A., Day,C.H.,
            Skeiky,Y.A. and Wang,A.
  TITLE     Compositions and methods for the therapy and diagnosis of prostate
            cancer
  JOURNAL   Patent: WO 0134802-A 127 17-MAY-2001;
            CORIXA CORPORATION (US)
FEATURES             Location/Qualifiers
     source          1. .54
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT       23 a     17 c      9 g      5 t
ORIGIN

  Query Match             2.6%;  Score 54;  DB 6;  Length 54;
  Best Local Similarity   100.0%;  Pred. No. 0.00047;
  Matches   54;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     724 CTGCTGTGCTTGGCAAGTGGATTGATGCTTCCATCTTGTTTGTGGTTTCGTGGT 777
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      54 CTGCTGTGCTTGGCAAGTGGATTGATGCTTCCATCTTGTTTGTGGTTTCGTGGT 1


RESULT 15
AX200497/c
LOCUS       AX200497                 54 bp    DNA     linear   PAT 29-AUG-2001
DEFINITION  Sequence 127 from Patent WO0151633.
ACCESSION   AX200497
VERSION     AX200497.1  GI:15390309
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 54)
  AUTHORS   Xu,J., Dillon,D.C., Mitcham,J.L., Harlocker,S.L., Jiang,Y.,
            Reed,S.G., Kalos,M.D., Fanger,G.R., Day,C.H., Retter,M.W.,
            Stolk,J.A., Skeiky,Y.A., Wang,A. and Meagher,M.J.
  TITLE     Compositions and methods for the therapy and diagnosis of prostate
            cancer
  JOURNAL   Patent: WO 0151633-A 127 19-JUL-2001;
            CORIXA CORPORATION (US)
FEATURES             Location/Qualifiers
     source          1. .54
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT       23 a     17 c      9 g      5 t
ORIGIN

  Query Match             2.6%;  Score 54;  DB 6;  Length 54;
  Best Local Similarity   100.0%;  Pred. No. 0.00047;
  Matches   54;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     724 CTGCTGTGCTTGGCAAGTGGATTGATGCTTCCATCTTGTTTGTGGTTTCGTGGT 777
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      54 CTGCTGTGCTTGGCAAGTGGATTGATGCTTCCATCTTGTTTGTGGTTTCGTGGT 1


Search completed: October 21, 2002, 20:20:04
Job time : 4661 secs
```

GenCore version 5.1.3

OM nucleic - nucleic search, using sw model

Run on: October 21, 2002, 17:33:38 ; Search time 342 Seconds
(without alignments)
10557.518 Million cell updates/sec

Title: US-09-809-638-1
Perfect score: 2103
Sequence: 1 atgacctcgctgtggagaga.....ccaaatacttttttatgaaac 2103

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_032802:*
1: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	384.8	18.3	635	23	AAS66555	DNA encoding novel
	2	142.8	6.8	670	23	AAS66554	DNA encoding novel
c	3	54	2.6	54	19	AAV61215	cDNA sequence of p
c	4	54	2.6	54	19	AAV58600	Prostate tumour sp
c	5	54	2.6	54	21	AAA06363	Human immunogenic
c	6	54	2.6	54	22	AAS63571	Human prostate cDN
c	7	54	2.6	54	22	AAS10122	Human prostate tum
c	8	54	2.6	54	22	AAH93479	Human prostate-spe
c	9	54	2.6	54	22	AAH84793	Human prostate-spe
c	10	54	2.6	54	22	AAH02544	Prostate tumour an
	11	38.4	1.8	7850	23	ABL24162	Drosophila melanog
c	12	38	1.8	2523	23	AAS72985	DNA encoding novel
c	13	38	1.8	9813	20	AAX13059	Enterococcus faeca
c	14	37.8	1.8	6533	24	ABL33708	Human immune syste
	15	37.2	1.8	616	21	AAZ80484	Human colon cancer
	16	37.2	1.8	6583	22	ABA19235	Human nervous syst
	17	37	1.8	236303	22	AAS11614	Human genomic DNA
c	18	36.6	1.7	6668	22	AAS46417	Tumour suppressor
c	19	36.6	1.7	6668	24	ABL33218	Human immune syste
c	20	36.4	1.7	32179	22	ABA17139	Human nervous syst
c	21	36.4	1.7	80331	22	AAC89559	Human histone deac
	22	36.2	1.7	8155	18	AAV74374	Staphylococcus aur
	23	36	1.7	417	22	AAS27442	cDNA encoding nove
	24	36	1.7	1734	22	AAH14130	Human cDNA sequenc
	25	36	1.7	2805	21	AAC81736	Human secreted pro
	26	36	1.7	4236	20	AAX84253	Human Nck associat
c	27	36	1.7	6018	22	AAS46427	Tumour suppressor
	28	36	1.7	14041	22	AAH48024	Internal control B
	29	35.8	1.7	1812	23	AAS67909	DNA encoding novel
c	30	35.8	1.7	5263	22	AAS46468	Tumour suppressor
c	31	35.8	1.7	5263	24	AAS61226	Human gene regulat
c	32	35.6	1.7	18283	24	AAS61363	Human gene regulat
c	33	35.6	1.7	56506	21	AAA69168	Bacteriophage Dp-1
c	34	35.6	1.7	151826	21	AAF22291	BAC containing rep
c	35	35.4	1.7	5276	24	ABL32150	Human immune syste
	36	35.4	1.7	9300	22	ABA15594	Human nervous syst
	37	35.4	1.7	127197	22	AAI61370	Soybean 515002 reg
	38	35.2	1.7	296	23	AAD23551	Human lung tumour-
c	39	35.2	1.7	846	23	AAS76980	DNA encoding novel
c	40	35.2	1.7	4015	22	AAH33952	Human colon cancer
c	41	35.2	1.7	4029	22	AAI58323	Human polynucleoti
	42	35.2	1.7	4331	22	AAI60109	Human polynucleoti
c	43	35.2	1.7	6197	24	ABL33710	Human immune syste
	44	35	1.7	415	22	AAK76562	Human immune/haema
	45	35	1.7	441	22	AAK76563	Human immune/haema

ALIGNMENTS

RESULT 1
AAS66555
ID AAS66555 standard; cDNA; 635 BP.
XX
AC AAS66555;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #2359.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR P-PSDB; ABG02368.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in

```
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity  -
XX
PS   Claim 1; SEQ ID No 2359; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human
CC   diagnostic coding sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 635 BP; 255 A; 107 C; 112 G; 160 T; 1 other;

  Query Match             18.3%;  Score 384.8;  DB 23;  Length 635;
  Best Local Similarity   98.8%;  Pred. No. 1.6e-105;
  Matches  408;  Conservative    0;  Mismatches    3;  Indels    2;  Gaps    2;

Qy   1693 GTTTCAAAACTACTGAAAAGTAGCTCTAATCAAGTGATATTTCTGGGATATATCACTTCA 1752
           ||||||||||||||| |||||||||||||||||||||||||||||||||||||||||||
Db      1 GTTTCAAAACTACTGNAAAGTAGCTCTAATCAAGTGATATTTCTGGGATATATCACTTCA 60

Qy   1753 GCACCTGGCTCCAGAGATTATCTACAGCTCACTGAACATGGCAATGTGAAGGATATCGAC 1812
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 GCACCTGGCTCCAGAGATTATCTACAGCTCACTGAACATGGCAATGTGAAGGATATCGAC 120

Qy   1813 AGCACTGATCATGACAGATGGTGTGAATACATTATGTATCGAGGGCTGATCAGGTTGGGT 1872
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 AGCACTGATCATGACAGATGGTGTGAATACATTATGTATCGAGGGCTGATCAGGTTGGGT 180

Qy   1873 TATGCAAGAATCT-CCCATGCTGAACTGAGTGATTCAGAAATTCAGATGGCAAAATTTAG 1931
           ||||||||||||| |||||||||||||||||||||||||||||||||||||||||||||
Db    181 TATGCAAGAATCTCCCCATGCTGAACTGAGTGATTCAGAAATTCAGATGGCAAAATTTAG 240

Qy   1932 GATCCCTGATGACCCCACTAATTATAGAGACAACCAGAAAGTGGTCATAGACCACAGAGA 1991
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 GATCCCTGATGACCCCACTAATTATAGAGACAACCAGAAAGTGGTCATAGACCACAGAGA 300

Qy   1992 AGTTTCTGAGAAAATTCATTTTAATCCCAGATTTGGATCCTACAAAGAAGGACACAATTA 2051
           |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db    301 AGTTTCTGAGAAAATTCATTTTAATCCCAGATTTGGATCCTACAAAGAAGGACCCAATTA 360

Qy   2052 TGAAAACAACCATCATTTTCATATGAATACT-CCCAAATACTTTTTATGAAAC 2103
           ||||||||||||| |||||||||||||||| |||||||||||||||||||||
Db    361 TGAAAACAACCATAATTTTCATATGAATACTCCCCAAATACTTTTTATGAAAC 413


RESULT 2
AAS66554
ID   AAS66554 standard; cDNA; 670 BP.
XX
AC   AAS66554;
XX
DT   13-FEB-2002  (first entry)
XX
DE   DNA encoding novel human diagnostic protein #2358.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss..
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   P-PSDB; ABG02367.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity  -
XX
PS   Claim 1; SEQ ID No 2358; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human
CC   diagnostic coding sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 670 BP; 181 A; 177 C; 132 G; 180 T; 0 other;

  Query Match              6.8%;  Score 142.8;  DB 23;  Length 670;
  Best Local Similarity   85.9%;  Pred. No. 2e-32;
  Matches  170;  Conservative    0;  Mismatches   27;  Indels    1;  Gaps    1;

Qy   1821 TCATGACAGATGGTGTGAATACATTATGTATCGAGGGCTGATCAGGTTGGGTTATGCAAG 1880
           ||| |  |     |||   |  | || |  |||     | |  ||||||||||||||||
Db    349 TCAGGTTATTACCTGTCTTTTAAATACGCTTCGGATAGTTACAAGGTTGGGTTATGCAAG 408

Qy   1881 AATCTCCCATGCTGAACTGAGTGATTCAGAAATTCAGATGGCAAAATTTAGGATCCCTGA 1940
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    409 AATCTCCCATGCTGAACTGAGTGATTCAGAAATTCAGATGGCAAAATTTAGGATCCCTGA 468

Qy   1941 TGACCCCACTAATTATAGAGACAACCAGAAAGTGGTCATAGACCACAGAGAAGTTTCTGA 2000
           |||||||||||||||||||||||||||||||| |||||||| ||||||||||||||||||
Db    469 TGACCCCACTAATTATAGAGACAACCAGAAAGTTGTCATAGA-CACAGAGAAGTTTCTGA 527

Qy   2001 GAAAATTCATTTTAATCC 2018
           |||||||||||||| ||
Db    528 GAAAATTCATTTTAACCC 545
```

RESULT 3
AAV61215/c
ID AAV61215 standard; cDNA; 54 BP.
XX
AC AAV61215;
XX
DT 06-JAN-1999 (first entry)
XX
DE cDNA sequence of prostate tumour clone.
XX
KW Prostate; cancer; tumour; vaccine; immunogen; clone; ss.
XX
OS Homo sapiens.
XX
PN WO9837093-A2.
XX
PD 27-AUG-1998.
XX
PF 25-FEB-1998; 98WO-US03492.
XX
PR 09-FEB-1998; 98US-0020956.
PR 25-FEB-1997; 97US-0806099.
PR 01-AUG-1997; 97US-0904804.
XX
PA (CORI-) CORIXA CORP.
XX
PI Dillon DC, Xu J;
XX
DR WPI; 1998-609886/51.
XX
PT Polypeptides comprising immunogenic portions of prostate proteins -
PT used in a vaccine for the treatment of prostate cancer
XX
PS Claim 3; Page 89; 130pp; English.
XX
CC The present sequence is a new DNA which encodes an immunogenic portion
CC of a prostate tumour protein. The encoded immunogen, or the DNA itself,
CC can be used as a vaccine for the treatment of prostate cancer. The DNA
CC was identified by analysis of a subtracted cDNA library obtained by
CC subtracting a prostate tumour cDNA expression library with a normal
CC tissue cDNA library.
XX
SQ Sequence 54 BP; 23 A; 17 C; 9 G; 5 T; 0 other;

Query Match 2.6%; Score 54; DB 19; Length 54;
Best Local Similarity 100.0%; Pred. No. 3.1e-06;
Matches 54; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 724 CTGCTGTGCTTGGCAAGTGGATTGATGCTTCCATCTTGTTTGTGGTTTCGTGGT 777
||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 54 CTGCTGTGCTTGGCAAGTGGATTGATGCTTCCATCTTGTTTGTGGTTTCGTGGT 1

RESULT 4
AAV58600/c
ID AAV58600 standard; cDNA; 54 BP.
XX
AC AAV58600;
XX
DT 08-DEC-1998 (first entry)
XX
DE Prostate tumour specific gene clone.
XX
KW Prostate tumour specific gene; human; prostate cancer; detection;
KW therapy; ss.
XX
OS Homo sapiens.
XX
PN WO9837418-A2.
XX
PD 27-AUG-1998.
XX
PF 25-FEB-1998; 98WO-US03690.
XX
PR 09-FEB-1998; 98US-0904809.
PR 25-FEB-1997; 97US-0806596.
PR 01-AUG-1997; 97US-0904809.
XX
PA (CORI-) CORIXA CORP.
XX
PI Dillon DC, Xu J;
XX
DR WPI; 1998-480805/41.
XX
PT Novel human prostate specific tumour protein and fragments - useful
PT for detecting and treating prostate cancers
XX
PS Claim 1; Page 95; 141pp; English.
XX
CC This sequence represents a human prostate tumour specific gene, and can
CC be used in the method of the invention. The method is for detecting
CC prostate cancer comprises contacting a biological sample with an agent
CC able to bind an immunogenic portion of a prostate protein (such as
CC encoded by this sequence). An antibody which binds to an immunogenic
CC portion of the prostate protein, and the method can be used to detect,
CC monitor progression of, or treat prostate cancers. The antibody may
CC also be conjugated to a therapeutic agent for use in therapy of prostate
CC cancers.
XX
SQ Sequence 54 BP; 23 A; 17 C; 9 G; 5 T; 0 other;

Query Match 2.6%; Score 54; DB 19; Length 54;
Best Local Similarity 100.0%; Pred. No. 3.1e-06;
Matches 54; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 724 CTGCTGTGCTTGGCAAGTGGATTGATGCTTCCATCTTGTTTGTGGTTTCGTGGT 777
||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 54 CTGCTGTGCTTGGCAAGTGGATTGATGCTTCCATCTTGTTTGTGGTTTCGTGGT 1

RESULT 5
AAA06363/c
ID AAA06363 standard; cDNA; 54 BP.
XX
AC AAA06363;
XX
DT 13-JUN-2000 (first entry)
XX
DE Human immunogenic prostate tumour protein cDNA sequence SEQ ID NO:127.
XX
KW Human; prostate cancer; diagnosis; tumour; gene therapy; detection;
KW immunogenic; cytostatic; vaccine; ss.
XX
OS Homo sapiens.
XX
PN WO200004149-A2.
XX
PD 27-JAN-2000.
XX
PF 14-JUL-1999; 99WO-US15838.
XX
PR 14-JUL-1998; 98US-0115453.
PR 14-JUL-1998; 98US-0116134.
PR 23-SEP-1998; 98US-0159812.
PR 23-SEP-1998; 98US-0159822.
PR 15-JAN-1999; 99US-0232149.
PR 15-JAN-1999; 99US-0232880.
PR 09-APR-1999; 99US-0288946.
XX
PA (CORI-) CORIXA CORP.
XX
PI Dillon DC, Harlocker SL, Yuqiu J, Xu J, Mitcham JL;
XX
DR WPI; 2000-171268/15.
XX
PT New polypeptide useful for treating and diagnosing prostate cancer

```
PT   comprises an immunogenic portion of prostate tumor protein -
XX
PS   Claim 1; Page 143; 263pp; English.
XX
CC   The present invention describes isolated polypeptides, comprising an
CC   immunogenic portion of a prostate tumour protein (PTP). The polypeptides
CC   and polynucleotides encoding them have cytostatic activity and can be
CC   used in vaccines and in gene therapy. The polypeptides and
CC   polynucleotides encoding them, antigen presenting cells which express
CC   the polypeptides, antibodies against the polypeptides and vaccines
CC   comprising them can be used for inhibiting the development of prostate
CC   cancer in a patient. The polypeptides can be used to generate antibodies
CC   or anti-idiotypic antibodies for passive immuno therapy. A portion of
CC   the polynucleotides encoding the polypeptides can be used as a probe or
CC   to modulate the expression of the polypeptides. AAA06241 to AAA06691 and
CC   AAY82000 to AAY82020 represent sequences used in the exemplification of
CC   the present invention.
XX
SQ   Sequence 54 BP; 23 A; 17 C; 9 G; 5 T; 0 other;

  Query Match             2.6%;  Score 54;  DB 21;  Length 54;
  Best Local Similarity   100.0%;  Pred. No. 3.1e-06;
  Matches   54;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     724 CTGCTGTGCTTGGCAAGTGGATTGATGCTTCCATCTTGTTTGTGGTTTCGTGGT 777
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      54 CTGCTGTGCTTGGCAAGTGGATTGATGCTTCCATCTTGTTTGTGGTTTCGTGGT 1


RESULT 6
AAS63571/c
ID   AAS63571 standard; cDNA; 54 BP.
XX
AC   AAS63571;
XX
DT   29-JAN-2002  (first entry)
XX
DE   Human prostate cDNA sequence #123.
XX
KW   Human; prostate cancer; ss; cytostatic; immunostimulant; tumour.
XX
OS   Homo sapiens.
XX
PN   WO200173032-A2.
XX
PD   04-OCT-2001.
XX
PF   27-MAR-2001; 2001WO-US09919.
XX
PR   27-MAR-2000; 2000US-0536857.
PR   09-MAY-2000; 2000US-0568100.
PR   12-MAY-2000; 2000US-0570737.
PR   13-JUN-2000; 2000US-0593793.
PR   27-JUN-2000; 2000US-0605783.
PR   10-AUG-2000; 2000US-0636215.
PR   29-AUG-2000; 2000US-0651236.
PR   06-SEP-2000; 2000US-0657279.
PR   02-OCT-2000; 2000US-0679426.
PR   10-OCT-2000; 2000US-0685166.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Xu J,  Dillon DC,  Mitcham JL,  Harlocker SL,  Jiang Y,  Kalos MD;
PI   Fanger GR,  Retter MW,  Stolk JA,  Day CH,  Vedvick TS,  Carter D;
PI   Li SX,  Wang A,  Skeiky YAW,  Hepler WT,  Henderson RA;
XX
DR   WPI; 2001-639232/73.
XX
PT   New human prostate-specific polypeptides and polynucleotides useful for
PT   the diagnosis and treatment of cancer, especially prostate cancer -
XX
PS   Claim 1; Page 274; 579pp; English.
XX
CC   The invention relates to isolated prostate-specific
CC   polynucleotides, polypeptides, fusion proteins of the polypeptides,
CC   antibodies raised against the polypeptides (or antigenic epitopes
CC   derived from them) and antigen-presenting cells expressing the
CC   polypeptides. The antibodies are useful for detecting the presence of
CC   cancer, especially prostate cancer. The polypeptides, polynucleotides and
CC   the antigen-presenting cells are useful for stimulating and/or expanding
CC   T cells specific for a tumour protein, and for inhibiting the development
CC   of cancer especially prostate cancer. Compositions comprising the
CC   polynucleotide and/or polypeptide are useful for stimulating an immune
CC   response, and for treating cancer. The oligonucleotide is useful for
CC   detecting cancer. The present sequence is a prostate specific
CC   polynucleotide of the invention.
XX
SQ   Sequence 54 BP; 23 A; 17 C; 9 G; 5 T; 0 other;

  Query Match             2.6%;  Score 54;  DB 22;  Length 54;
  Best Local Similarity   100.0%;  Pred. No. 3.1e-06;
  Matches   54;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     724 CTGCTGTGCTTGGCAAGTGGATTGATGCTTCCATCTTGTTTGTGGTTTCGTGGT 777
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      54 CTGCTGTGCTTGGCAAGTGGATTGATGCTTCCATCTTGTTTGTGGTTTCGTGGT 1


RESULT 7
AAS10122/c
ID   AAS10122 standard; cDNA; 54 BP.
XX
AC   AAS10122;
XX
DT   24-OCT-2001  (first entry)
XX
DE   Human prostate tumour cDNA #13.
XX
KW   Human; prostate tumour protein; prostate cancer; ss.
XX
OS   Homo sapiens.
XX
PN   US6262245-B1.
XX
PD   17-JUL-2001.
XX
PF   25-FEB-1998;   98US-0030607.
XX
PR   25-FEB-1997;   97US-0806099.
PR   01-AUG-1997;   97US-0904804.
PR   09-FEB-1998;   98US-0020956.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Xu J,  Dillon DC;
XX
DR   WPI; 2001-440862/47.
XX
PT   Novel polynucleotide encoding polypeptide comprising a portion of
PT   prostate tumour protein useful for inhibiting development of prostate
PT   cancer or for treating prostate cancer in a patient  -
XX
PS   Example 2; Column 137; 105pp; English.
XX
CC   The sequence is a human prostate tumour cDNA which encodes a
CC   partial tumour protein. The DNA is useful for inhibiting the development
CC   of prostate cancer or for treating prostate cancer in a patient.
XX
SQ   Sequence 54 BP; 23 A; 17 C; 9 G; 5 T; 0 other;

  Query Match             2.6%;  Score 54;  DB 22;  Length 54;
  Best Local Similarity   100.0%;  Pred. No. 3.1e-06;
  Matches   54;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     724 CTGCTGTGCTTGGCAAGTGGATTGATGCTTCCATCTTGTTTGTGGTTTCGTGGT 777
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db     54 CTGCTGTGCTTGGCAAGTGGATTGATGCTTCCATCTTGTTTGTGGTTTCGTGGT 1


RESULT 8
AAH93479/c
ID   AAH93479 standard; cDNA; 54 BP.
XX
AC   AAH93479;
XX
DT   04-OCT-2001  (first entry)
XX
DE   Human prostate-specific cDNA sequence P126.
XX
KW   Human; prostate cancer; prostate-specific; diagnosis; vaccine;
KW   cytostatic; gene therapy; metastasis; ss.
XX
OS   Homo sapiens.
XX
PN   WO200151633-A2.
XX
PD   19-JUL-2001.
XX
PF   16-JAN-2001; 2001WO-US01574.
XX
PR   14-JAN-2000; 2000US-0483672.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Xu J, Dillon DC, Mitcham JL, Harlocker SL, Jiang Y, Reed SG;
PI   Kalos MD, Fanger GR, Day CH, Retter MW, Stolk JA, Skeiky YAW;
PI   Wang A, Meagher MJ;
XX
DR   WPI; 2001-425873/45.
XX
PT   New polynucleotide encoding a prostate-specific protein, for
PT   diagnosing, monitoring and treating prostate cancer in a patient and
PT   for use in vaccines -
XX
PS   Claim 1; Page 272; 543pp; English.
XX
CC   The present invention describes polynucleotide sequences (I) which encode
CC   prostate-specific proteins (II). (I) and (II) have cytostatic activity,
CC   and can be used in vaccine production and gene therapy. (I), (II),
CC   antibodies to (II), fusion proteins comprising (II), and isolated
CC   T cells prepared using (I) or (II) are used treat cancer in a patient.
CC   (I) and the antibodies are also used in the detection of cancer in a
CC   patient. The cancer that is diagnosed or treated is particularly
CC   prostate cancer. (I) and (II) can be used in vaccines. The antibodies or
CC   (I) can be used for monitoring the progression of cancer in a patient.
CC   (I) and (II) can also be used to improve diagnostic and therapeutic
CC   methods for prostate cancer. They can indicate the level of metastasis
CC   as well as the prostate volume. AAH93357 to AAH93944 and AAM01115 to
CC   AAM01318 represent polynucleotide and amino acid sequences used in the
CC   exemplification of the present invention.
XX
SQ   Sequence 54 BP; 23 A; 17 C; 9 G; 5 T; 0 other;

  Query Match             2.6%;  Score 54;  DB 22;  Length 54;
  Best Local Similarity   100.0%;  Pred. No. 3.1e-06;
  Matches   54;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    724 CTGCTGTGCTTGGCAAGTGGATTGATGCTTCCATCTTGTTTGTGGTTTCGTGGT 777
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     54 CTGCTGTGCTTGGCAAGTGGATTGATGCTTCCATCTTGTTTGTGGTTTCGTGGT 1


RESULT 9
AAH84793/c
ID   AAH84793 standard; cDNA; 54 BP.
XX
AC   AAH84793;
XX
DT   25-SEP-2001  (first entry)
XX
DE   Human prostate-specific cDNA sequence P126.
XX
KW   Human; prostate cancer; therapy; diagnosis; cat eye syndrome;
KW   chromosome 22q11.2; prostate-specific protein; chromosome 1;
KW   prostate specific antigen; PSA; ss.
XX
OS   Homo sapiens.
XX
PN   WO200134802-A2.
XX
PD   17-MAY-2001.
XX
PF   09-NOV-2000; 2000WO-US30904.
XX
PR   12-NOV-1999;   99US-0439313.
PR   18-NOV-1999;   99US-0443686.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Xu J, Dillon DC, Mitcham JL, Harlocker SL, Jiang Y, Reed SG;
PI   Kalos MD, Retter MW, Stolk JA, Day CH, Skeiky YAW, Wang A;
XX
DR   WPI; 2001-308785/32.
XX
PT   Isolated polypeptide comprising at least an immunogenic portion of a
PT   prostate-specific protein, useful in the diagnosis and therapy of
PT   prostate cancer -
XX
PS   Claim 5; Page 172; 325pp; English.
XX
CC   The present invention describes an isolated polypeptide (P1) comprising
CC   at least an immunogenic portion of a prostate-specific protein, or its
CC   variant. Also described are polynucleotides (N1) encoding (P1). (P1) and
CC   (N1) have cytostatic activity and can be used in vaccine production.
CC   The polypeptides, nucleic acids and antibodies from the present
CC   invention are useful in the diagnosis and therapy of prostate cancer.
CC   Prostate specific genes P704P, P712P, P774P, P775P and B305D are located
CC   in a genomic region on chromosome 22q11.2 known as the Cat Eye Syndrome
CC   region. Prostate specific antigen (PSA) P501S was located on
CC   chromosome 1. AAH84671 to AAH85143 and AAG99000 to AAG99077 represent
CC   polynucleotide and polypeptide sequences used in the exemplification
CC   of the present invention.
XX
SQ   Sequence 54 BP; 23 A; 17 C; 9 G; 5 T; 0 other;

  Query Match             2.6%;  Score 54;  DB 22;  Length 54;
  Best Local Similarity   100.0%;  Pred. No. 3.1e-06;
  Matches   54;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    724 CTGCTGTGCTTGGCAAGTGGATTGATGCTTCCATCTTGTTTGTGGTTTCGTGGT 777
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     54 CTGCTGTGCTTGGCAAGTGGATTGATGCTTCCATCTTGTTTGTGGTTTCGTGGT 1


RESULT 10
AAH02544/c
ID   AAH02544 standard; cDNA; 54 BP.
XX
AC   AAH02544;
XX
DT   14-JUN-2001  (first entry)
XX
DE   Prostate tumour antigen determined cDNA sequence for P126.
XX
KW   Human; prostate tumour antigen; prostate tumour; therapy; diagnosis;
KW   prostate cancer; immunogenic; cytostatic; vaccine; ss.
XX
OS   Homo sapiens.
XX
PN   WO200125272-A2.
XX
PD   12-APR-2001.
```

XX
PF 04-OCT-2000; 2000WO-US27464.
XX
PR 04-OCT-1999; 99US-0157455.
XX
PA (CORI-) CORIXA CORP.
XX
PI Xu J, Skeiky YAW, Reed SG, Cheever MA;
XX
DR WPI; 2001-245062/25.
XX
PT Prostate specific protein and its encoding polynucleotide, useful for
PT the treatment and diagnosis of prostate cancer -
XX
PS Claim 4; Page 162; 276pp; English.
XX
CC The present invention describes an isolated polypeptide (I) comprising
CC at least an immunogenic portion of a prostate tumour antigen protein or
CC its variant. (I) have cytostatic activity and can be used in vaccine
CC production. (I), prostate tumour antigen polynucleotides, an antigen
CC presenting cell (APC e.g. a dendritic cell) that expresses (I), and a
CC pharmaceutical composition containing (I) are useful for inhibiting the
CC development of cancer in a patient. Antibodies specific for prostate
CC specific proteins and oligonucleotides that hybridise to a
CC polynucleotide that encodes a prostate specific protein are useful
CC for detecting the presence or absence of a cancer or monitoring the
CC progression the progression of a cancer, especially prostate cancer.
CC AAH02422 to AAH2872, AAB74798 to AAB74821 and AAB74830 are sequences
CC used in the exemplification of the present invention.
XX
SQ Sequence 54 BP; 23 A; 17 C; 9 G; 5 T; 0 other;

Query Match 2.6%; Score 54; DB 22; Length 54;
Best Local Similarity 100.0%; Pred. No. 3.1e-06;
Matches 54; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 724 CTGCTGTGCTTGGCAAGTGGATTGATGCTTCCATCTTGTTTGTGGTTTCGTGGT 777
||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 54 CTGCTGTGCTTGGCAAGTGGATTGATGCTTCCATCTTGTTTGTGGTTTCGTGGT 1

RESULT 11
ABL24162
ID ABL24162 standard; DNA; 7850 BP.
XX
AC ABL24162;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 23959.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ds.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Claim 1; SEQ ID NO 23959; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 7850 BP; 1994 A; 1770 C; 1831 G; 2255 T; 0 other;

Query Match 1.8%; Score 38.4; DB 23; Length 7850;
Best Local Similarity 50.0%; Pred. No. 2.8;
Matches 96; Conservative 0; Mismatches 96; Indels 0; Gaps 0;

Qy 1070 TGTTAATTATCGGGCTGAATATGCTATTTGGTCCTAAGAAAAACCTTGACTTGCTTCTTC 1129
|| |||| || ||||| |||| |||| | || | ||| | ||||
Db 5814 TGCTAATCCGCGTGCTGATTATGATATTCGTACCCTAAGGTGTCCTAAATTTGCCCAGAA 5873

Qy 1130 AAACAAAAAAACAGTTCTAAAGTGCTTTTCAGAAAGAGTGAAAAATACATGAAACTTTTTC 1189
||| ||| | | ||| || || ||| | ||| | || || ||
Db 5874 GTTTCCAAATCAGATAAGCGGAGGTTTGAAGTAAATGTGCATAATCCTTGCAAATTAGCA 5933

Qy 1190 TGTGGCTGCTTGTTGGTGTGGGATTGTTGGGATTAGGACTACGGCATAAAGCCTATGAGA 1249
|| | || | | | || | || | || | ||| | || |
Db 5934 CGTAGATGGTAGCACTTCTTTGAGTATTAAGTAATGGCTTTGTTCATCTATAGACTGCTA 5993

Qy 1250 GAAAACTGGGCA 1261
|||| || ||
Db 5994 GAAATGTGTTCA 6005

RESULT 12
AAS72985/c
ID AAS72985 standard; cDNA; 2523 BP.
XX
AC AAS72985;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #8789.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR P-PSDB; ABG08798.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess

PT biodiversity -
XX
PS Claim 1; SEQ ID No 8789; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 2523 BP; 526 A; 803 C; 669 G; 525 T; 0 other;

Query Match 1.8%; Score 38; DB 23; Length 2523;
Best Local Similarity 49.0%; Pred. No. 1.9;
Matches 101; Conservative 0; Mismatches 105; Indels 0; Gaps 0;

```
Qy    640 GGAGAAGTCTCTCTTGTTTCCAGATGGGCAGTGAGTGGGCATCCACATCCAGGGCCAGAT 699
          |||| ||  | |||    |||||||   | | |||  ||  ||  ||| ||  | ||  |
Db    598 GGAGCAGAGTTTCTGACCTCCAGATCTCCTGGGAGGCGGGGTCTGCATTCACAGACATCT 539

Qy    700 CCTAACCCATTTGGAGGTGCAGTACTGCTGTGCTTGGCAAGTGGATTGATGCTTCCATCT 759
            | |   |  ||  | | |||       |||    | | |      || ||      |
Db    538 GATGAGAGAAGTGTCGCTTCAGGGGGCTAGTGTGCTGGAGGCTACAGGAAGCAAAGGGCA 479

Qy    760 TGTTTGTGGTTTCGTGGTACTGGTTTGATCTGGTGGGTTACAGGAACAGCTTCAGCTGCG 819
          ||  ||  || |    |   ||||  || || ||| || ||    |      | ||| |
Db    478 TGGCTGGTGTCTTCCTGGTGTGGTGGCATGTGCTGGCTTGCACATGCCAGCAGATCTGGG 419

Qy    820 GGGCTCCTTTACCTGCACACATGGGC 845
          | |  ||   | ||||| | | ||||
Db    418 GTGGGCCAGCATCTGCAGAAACGGGC 393
```

RESULT 13
AAX13059/c
ID AAX13059 standard; DNA; 9813 BP.
XX
AC AAX13059;
XX
DT 19-MAR-1999 (first entry)
XX
DE Enterococcus faecalis genome contig SEQ ID NO:122.
XX
KW Enterococcus faecalis; contig; detection; Enterococcal infection;
KW vaccine; attenuation; computer readable medium; ds.
XX
OS Enterococcus faecalis.
XX
PN WO9850555-A2.
XX
PD 12-NOV-1998.
XX
PF 04-MAY-1998; 98WO-US08985.
XX
PR 14-NOV-1997; 97US-0066009.
PR 06-MAY-1997; 97US-0044031.
PR 16-MAY-1997; 97US-0046655.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Barash SC, Dillon PJ, Kunsch CA;
XX
DR WPI; 1999-045171/04.
XX
PT New isolated Enterococcus faecalis polynucleotides and polypeptides
PT - used to develop products for the detection of Enterococcus and for
PT use in vaccines for prevention or attenuation of Enterococcus
PT infection.
XX
PS Claim 1; Page 734-739; 2084pp; English.
XX
CC A computer readable medium has been developed which has recorded on it
CC 982 nucleotide sequences isolated from the Enterococcus faecalis genome.
CC AAX12938 to AAX13919 represent these nucleotide sequences which are
CC primary nucleotide sequences, also known as contigs. The computer-based
CC system can identify fragments of the Enterococcus faecalis genome with
CC commercial importance. The products can be used to detect the presence
CC of Enterococcus faecalis in samples. They can also be used for
CC diagnosing Enterococcal infection in an animal and monitoring
CC progression of disease, and for identifying agents which can be used to
CC modulate the growth or pathogenicity of Enterococcus faecalis, or
CC another related organism, in vivo or in vitro. In particular the
CC polypeptides encoded by the Enterococcus faecalis nucleotide sequences
CC can be used in vaccines to prevent or attenuate an Enterococcal
CC infection.
XX
SQ Sequence 9813 BP; 3233 A; 1698 C; 2161 G; 2719 T; 2 other;

Query Match 1.8%; Score 38; DB 20; Length 9813;
Best Local Similarity 49.0%; Pred. No. 4.1;
Matches 101; Conservative 0; Mismatches 105; Indels 0; Gaps 0;

```
Qy   1106 AGAAAAACCTTGACTTGCTTCTTCAAACAAAAAACAGTTCTAAAGTGCTTTTCAGAAAGA 1165
          | |||  |   || ||  |||    |||  |||| |||||   | | ||||| | ||| |
Db   2747 AAAAATTCGAGGATTTTTTTCAAATAACTGAAAATAGTTCGCCATTTCTTTTAAAAAAAA 2688

Qy   1166 GTGAAAAATACATGAAACTTTTTCTGTGGCTGCTTGTTGGTGTGGGATTGTTGGGATTAG 1225
          |  | |||      |||||||   | |      | |||  | ||   || |||   |
Db   2687 GAAATAAAGGAGAAAAACTTTCGTTTTTCTCCTTAGTTTCTTTGCATTTTTTGTTTTCGT 2628

Qy   1226 GACTACGGCATAAAGCCTATGAGAGAAAACTGGGCAAAGTGGCACCAACCAAAGAGGTCT 1285
               |   | ||   ||  | || |||  | | | ||   | |  ||||||  |    |
Db   2627 TGGACCATGACAACCACTCGGCGATAAATATTGACTAATAAGAATAAACCAATAAAAATT 2568

Qy   1286 CTGCTGCCATCTGGCCTTTCAGGTTT 1311
            | |   |  ||  ||    ||| |
Db   2567 AGGGTAATACTTGCACTCGGTGGTGT 2542
```

RESULT 14
ABL33708/c
ID ABL33708 standard; DNA; 6533 BP.
XX
AC ABL33708;
XX
DT 26-MAR-2002 (first entry)
XX
DE Human immune system associated gene SEQ ID NO: 1681.
XX
KW Human; immune system disease; cytosine methylation; antiasthmatic;
KW antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
KW gene; ds.
XX
OS Homo sapiens.
XX

```
PN   WO200200928-A2.
XX
PD   03-JAN-2002.
XX
PF   02-JUL-2001; 2001WO-EP07537.
XX
PR   30-JUN-2000; 2000DE-1032529.
PR   01-SEP-2000; 2000DE-1043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2002-130909/17.
XX
PT   Nucleic acid comprising fragment of chemically modified gene, useful
PT   for diagnosis and treatment of diseases associated with abnormal
PT   cytosine methylation  -
XX
PS   Claim 1; SEQ ID NO 1681; 32pp + Sequence Listing; German.
XX
CC   The present invention provides a number of human immune system associated
CC   genes which are modified by the methylation of cytosines. The sequences
CC   can be used in the diagnosis and treatment of immune system disorders,
CC   including eye diseases such as retinopathy, neovascular glaucoma and
CC   macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC   leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC   rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC   diseases. The present sequence is a gene of the invention.
XX
SQ   Sequence 6533 BP; 1675 A; 132 C; 1539 G; 3187 T; 0 other;

  Query Match             1.8%;  Score 37.8;  DB 24;  Length 6533;
  Best Local Similarity   49.7%;  Pred. No. 3.8;
  Matches   96;  Conservative    0;  Mismatches   97;  Indels    0;  Gaps    0;

Qy    1910 AAATTCAGATGGCAAAATTTAGGATCCCTGATGACCCCACTAATTATAGAGACAACCAGA 1969
           ||| |   ||   |||   ||  ||| |  |  | | | |  |||||      | | | |
Db    6224 AAACTACAATAAAAAATCATACTATCACCAAACAACACGCACATTATTTTTTAACCGAAA 6165

Qy    1970 AAGTGGTCATAGACCACAGAGAAGTTTCTGAGAAAATTCATTTTAATCCCAGATTTGGAT 2029
           ||     || | |   || | ||   ||  | |||    ||   ||   || |     |
Db    6164 AACAAAACAAACATACCAAATAAAACTCCCAAAAATAAAATCAAAAAATCAAAACAAAAA 6105

Qy    2030 CCTACAAAGAAGGACACAATTATGAAAACAACCATCATTTTCATATGAATACTCCCAAAT 2089
           || ||||| ||  |   ||| ||  ||| |||||  ||||    |   | || | |  |
Db    6104 CCAACAAAAAAATAACAAATAATATAAAAAACCAAAATTTCTTAAAACACACGCACTTAA 6045

Qy    2090 ACTTTTTATGAAA 2102
           | ||| | | | |
Db    6044 AATTTATTTAATA 6032

RESULT 15
AAZ80484
ID   AAZ80484 standard; cDNA; 616 BP.
XX
AC   AAZ80484;
XX
DT   07-APR-2000  (first entry)
XX
DE   Human colon cancer cell line SW480 cDNA clone SEQ ID NO:568.
XX
KW   Human; gene expression product; diagnosis; tumour; colon cancer;
KW   colorectal adenocarcinoma; cell line SW480; cell proliferation;
KW   cytostatic; sarcoma; breast cancer; neoplasia; dysplasia;
KW   hyperplasia; ds.
XX
OS   Homo sapiens.
XX
PN   WO9964576-A2.
XX
PD   16-DEC-1999.
XX
PF   09-JUN-1999;   99WO-IB01062.
XX
PR   10-JUN-1998;   98US-0088801.
XX
PA   (FARB ) BAYER CORP.
XX
PI   Endege WO,  Steinmann KE,  Astle JH,  Burgess CC,  Bushnell SE;
PI   Carroll E,  Catino TJ,  Derti A,  Ford DM,  Lewis ME,  Monahan JE;
PI   Schlegel R;
XX
DR   WPI; 2000-087220/07.
XX
PT   Novel nucleic acids, used to develop products for the diagnosis and
PT   treatment of disorders involving unwanted cell proliferation,
PT   particularly cancers, especially colon cancer  -
XX
PS   Claim 15; Page 356; 469pp; English.
XX
CC   AAZ79917 to AAZ80766 represent double stranded cDNA clones isolated from
CC   the human colorectal adenocarcinoma (colon cancer) cell line SW480. The
CC   cDNA clones can be used to generate antisense oligonucleotides which
CC   can be used for antisense therapy. Methods and products from the present
CC   invention can be used for identifying and/or classifying cancerous cells
CC   present in a human tumour, particularly in solid tumours, e.g.
CC   carcinomas and sarcomas, e.g. breast or colon cancers. The cDNA clones
CC   can be used for developing agents for the diagnosis and treatment of
CC   disorders involving unwanted cell proliferation, such as neoplasia,
CC   dysplasia or hyperplasia.
XX
SQ   Sequence 616 BP; 145 A; 142 C; 163 G; 144 T; 22 other;

  Query Match             1.8%;  Score 37.2;  DB 21;  Length 616;
  Best Local Similarity   58.3%;  Pred. No. 1.5;
  Matches   63;  Conservative    0;  Mismatches   45;  Indels    0;  Gaps    0;

Qy    1106 AGAAAAACCTTGACTTGCTTCTTCAAACAAAAAACAGTTCTAAAGTGCTTTTCAGAAAGA 1165
           ||| | | || |  |  | ||||  |||||||  ||  | |    |  | ||||||||||
Db     290 AGACAGATCTGGTTTACCATCTTTTAACAAAAGGCAAATGTCTTCTCTTCTTCAGAAAGA 349

Qy    1166 GTGAAAAATACATGAAACTTTTTCTGTGGCTGCTTGTTGGTGTGGGAT 1213
           || |  || ||   ||   ||||||   |  | || ||  | |  |||
Db     350 GTCATTAACACTAAAATTCTTTTCTTNNGAAGTTTCTTCTTTTCCGAT 397


Search completed: October 21, 2002, 19:03:07
Job time : 400 secs
```

GenCore version 5.1.3

OM nucleic - nucleic search, using sw model

Run on: October 21, 2002, 18:56:23 ; Search time 69 Seconds
(without alignments)
7486.482 Million cell updates/sec

Title: US-09-809-638-1
Perfect score: 2103
Sequence: 1 atgacctcgctgtggagaga.....ccaaatacttttttatgaaac 2103

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

[redacted]rched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
[redacted]	1	54	2.6	54	4	US-09-020-956-127	Sequence 127, App
c	2	54	2.6	54	4	US-09-030-607-127	Sequence 127, App
c	3	54	2.6	54	4	US-09-439-313-127	Sequence 127, App
c	4	38.2	1.8	7218	1	US-08-232-463-14	Sequence 14, Appl
	5	37.2	1.8	616	4	US-09-328-111-568	Sequence 568, App
	6	37.2	1.8	7218	1	US-08-232-463-14	Sequence 14, Appl
c	7	34.8	1.7	1858	2	US-08-359-705B-7	Sequence 7, Appli
c	8	34.8	1.7	1858	2	US-08-286-846A-7	Sequence 7, Appli
c	9	34.8	1.7	1858	2	US-08-457-880A-7	Sequence 7, Appli
c	10	34.8	1.7	1858	3	US-08-444-622A-7	Sequence 7, Appli
c	11	34.8	1.7	1858	3	US-08-942-562-7	Sequence 7, Appli
c	12	34.8	1.7	1858	3	US-09-156-923-7	Sequence 7, Appli
c	13	34.8	1.7	2715	2	US-08-359-705B-5	Sequence 5, Appli
c	14	34.8	1.7	2715	2	US-08-286-846A-5	Sequence 5, Appli
c	15	34.8	1.7	2715	2	US-08-457-880A-5	Sequence 5, Appli
c	16	34.8	1.7	2715	3	US-08-444-622A-5	Sequence 5, Appli
c	17	34.8	1.7	2715	3	US-08-942-562-5	Sequence 5, Appli
c	18	34.8	1.7	2715	3	US-09-156-923-5	Sequence 5, Appli
c	19	34.8	1.7	2940	1	US-08-286-305A-8	Sequence 8, Appli
c	20	34.8	1.7	2940	2	US-08-441-104A-8	Sequence 8, Appli
c	21	34.8	1.7	2940	2	US-08-440-816A-8	Sequence 8, Appli
c	22	34.8	1.7	2940	4	US-09-417-381A-8	Sequence 8, Appli
c	23	33.6	1.6	1694	4	US-09-362-473-3	Sequence 3, Appli
	24	33.6	1.6	2094	4	US-09-106-194-1	Sequence 1, Appli
c	25	32.8	1.6	1683	4	US-09-009-443-6	Sequence 6, Appli
	26	32	1.5	2310	1	US-08-416-581B-2	Sequence 2, Appli
	27	32	1.5	2787	1	US-08-416-581B-3	Sequence 3, Appli
	28	32	1.5	2787	1	US-08-416-581B-4	Sequence 4, Appli
	29	32	1.5	2787	3	US-09-288-461-1	Sequence 1, Appli
	30	32	1.5	2787	4	US-09-087-465-5	Sequence 5, Appli
	31	31.8	1.5	1480	1	US-08-484-105-19	Sequence 19, Appl
	32	31.8	1.5	1480	1	US-08-484-106-19	Sequence 19, Appl
	33	31.8	1.5	3437	3	US-08-860-339-17	Sequence 17, Appl
c	34	31.6	1.5	151	4	US-09-439-313-319	Sequence 319, App
c	35	31.4	1.5	533	4	US-09-020-956-71	Sequence 71, Appl
c	36	31.4	1.5	533	4	US-09-030-607-71	Sequence 71, Appl
c	37	31.4	1.5	533	4	US-09-439-313-71	Sequence 71, Appl
	38	31	1.5	1667	1	US-08-485-284A-1	Sequence 1, Appli
	39	30.8	1.5	580	4	US-09-228-986-23	Sequence 23, Appl
c	40	30.8	1.5	2690	4	US-09-029-755C-1	Sequence 1, Appli
c	41	30.8	1.5	3539	4	US-09-508-542-16	Sequence 16, Appl
	42	30.8	1.5	19011	1	US-08-310-356-36	Sequence 36, Appl
	43	30.8	1.5	19557	5	PCT-US92-06300-1	Sequence 1, Appli
	44	30.6	1.5	1635	3	US-08-974-180-16	Sequence 16, Appl
	45	30.6	1.5	2178	3	US-08-974-180-11	Sequence 11, Appl

ALIGNMENTS

RESULT 1
US-09-020-956-127/c
; Sequence 127, Application US/09020956
; Patent No. 6261562
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillin, Davin C.
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY OF PROSTATE CANCER AND METHODS
; NUMBER OF SEQUENCES: 178
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/020,956
; FILING DATE: 09-FEB-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.427C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 127:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 54 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
US-09-020-956-127

Query Match 2.6%; Score 54; DB 4; Length 54;
Best Local Similarity 100.0%; Pred. No. 3.8e-07;
Matches 54; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 724 CTGCTGTGCTTGGCAAGTGGATTGATGCTTCCATCTTGTTTGTGGTTTCGTGGT 777
||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 54 CTGCTGTGCTTGGCAAGTGGATTGATGCTTCCATCTTGTTTGTGGTTTCGTGGT 1

```
RESULT 2
US-09-030-607-127/c
; Sequence 127, Application US/09030607
; Patent No. 6262245
;  GENERAL INFORMATION:
;    APPLICANT:  Xu, Jiangchun
;    APPLICANT:  Dillon, Davin C.
;    TITLE OF INVENTION:  COMPOUNDS FOR IMMUNOTHERAPY OF PROSTATE CANCER AND METHODS FO
;    NUMBER OF SEQUENCES:  224
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED and BERRY LLP
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  WA
;      COUNTRY:  USA
;      ZIP:  98104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/030,607
;      FILING DATE:  25-FEB-1998
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Maki, David J.
;      REGISTRATION NUMBER:  31,392
;      REFERENCE/DOCKET NUMBER:  210121.427C3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;  INFORMATION FOR SEQ ID NO:  127:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  54 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
US-09-030-607-127

  Query Match             2.6%;  Score 54;  DB 4;  Length 54;
  Best Local Similarity   100.0%;  Pred. No. 3.8e-07;
  Matches   54;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     724 CTGCTGTGCTTGGCAAGTGGATTGATGCTTCCATCTTGTTTGTGGTTTCGTGGT 777
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      54 CTGCTGTGCTTGGCAAGTGGATTGATGCTTCCATCTTGTTTGTGGTTTCGTGGT 1


RESULT 3
US-09-439-313-127/c
; Sequence 127, Application US/09439313
; Patent No. 6329505
; GENERAL INFORMATION:
;  APPLICANT: Xu, Jiangchun
;  APPLICANT:  Dillon, Davin C.
;  APPLICANT:  Mitcham, Jennifer L.
;  APPLICANT:  Harlocker, Susan Louise
;  APPLICANT:  Jiang Yuqui
;  APPLICANT:  Reed, Steven G.
;  APPLICANT:  Kalos, Michael
;  APPLICANT:  Fanger, Gary
;  APPLICANT:  Retter, Mark
;  APPLICANT:  Solk, John
;  APPLICANT:  Day, Craig
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
;  TITLE OF INVENTION:  DIAGNOSIS OF PROSTATE CANCER
;  FILE REFERENCE: 210121.427C9
;  CURRENT APPLICATION NUMBER: US/09/439,313
;  CURRENT FILING DATE:  1999-11-12
;  NUMBER OF SEQ ID NOS: 575
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 127
;   LENGTH: 54
;   TYPE: DNA
;   ORGANISM: Homo sapien
US-09-439-313-127

  Query Match             2.6%;  Score 54;  DB 4;  Length 54;
  Best Local Similarity   100.0%;  Pred. No. 3.8e-07;
  Matches   54;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     724 CTGCTGTGCTTGGCAAGTGGATTGATGCTTCCATCTTGTTTGTGGTTTCGTGGT 777
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      54 CTGCTGTGCTTGGCAAGTGGATTGATGCTTCCATCTTGTTTGTGGTTTCGTGGT 1


RESULT 4
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
;  GENERAL INFORMATION:
;    APPLICANT:  DORNER, F.
;    APPLICANT:  SCHEIFLINGER, F.
;    APPLICANT:  FALKNER, F. G.
;    TITLE OF INVENTION:  RECOMBINANT FOWLPOX VIRUS
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  1800 Diagonal Road, Suite 500
;      CITY:  Alexandria
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22313-0299
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/232,463
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/935,313
;      FILING DATE:
;      APPLICATION NUMBER:  EP 91 114 300.6
;      FILING DATE:  26-AUG-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BENT, Stephen A.
;      REGISTRATION NUMBER:  29,768
;      REFERENCE/DOCKET NUMBER:  30472/114 IMMU
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)836-9300
;      TELEFAX:  (703)683-4109
;      TELEX:  899149
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  7218 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      CLONE:  pTZgpt-Fls
US-08-232-463-14

  Query Match             1.8%;  Score 38.2;  DB 1;  Length 7218;
  Best Local Similarity   10.0%;  Pred. No. 0.34;
  Matches   46;  Conservative  200;  Mismatches  213;  Indels    0;  Gaps    0;

Qy    1126 CTTCAAACAAAAAACAGTTCTAAAGTGCTTTTCAGAAAGAGTGAAAAATACATGAAACTT 1185
```

```
Db    1490 CATCACTGTAATTACCTATCTATGCAAGTAGTTAAAGAGATAGAAGAATTTGGTACRRRR 1431

Qy    1186 TTTCTGTGGCTGCTTGTTGGTGTGGGATTGTTGGGATTAGGACTACGGCATAAAGCCTAT 1245

Db    1430 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1371

Qy    1246 GAGAGAAAACTGGGCAAAGTGGCACCAACCAAAGAGGTCTCTGCTGCCATCTGGCCTTTC 1305

Db    1370 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1311

Qy    1306 AGGTTTGGATATGACAATGAAGGGTGGTCTAGTCTAGAAAGATCAGCTCACCTGCTCAAT 1365

Db    1310 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1251

Qy    1366 GAAACAGGTGCAGATTTCATAACAATTTTGGAGAGTGATGCTTCTAAGCCCTATATGGGG 1425

Db    1250 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1191

Qy    1426 AACAATGACTTAACCATGTGGCTAGGGGAAAAGTTGGGTTTCTATACAGACTTTGGTCCA 1485

Db    1190 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1131

Qy    1486 AGCACAAGGTATCACACTTGGGGGATTATGGCTTTGTCAAGATACCCAATTGTGAAATCT 1545

Db    1130 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1071

Qy    1546 GAGCATCACCTTCTTCCGTCACCAGAGGGCGAGATCGCA 1584

Db    1070 RRRRATCGCAAGCTCCCTCGACCTGCAGCCAAGCTCGGA 1032


RESULT 5
US-09-328-111-568
; Sequence 568, Application US/09328111
; Patent No. 6262333
; GENERAL INFORMATION:
;  APPLICANT: Endege, Wilson O.
;  APPLICANT: Steinmann, Kathleen E.
;  APPLICANT: Astle, Jon H.
;  APPLICANT: Burgess, Christopher C.
;  APPLICANT: Bushnell, Steven E.
;  APPLICANT: Carroll III, Eddie
;  APPLICANT: Catino, Theodore J.
;  APPLICANT: Derti, Adnan
;  APPLICANT: Ford, Donna M.
;  APPLICANT: Lewis, Marcia E.
;  APPLICANT: Monahan, John E.
;  APPLICANT: Schlegel, Robert
;  TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
;  TITLE OF INVENTION: PRODUCTS
;  FILE REFERENCE: CCD-257 (US)
;  CURRENT APPLICATION NUMBER: US/09/328,111
;  CURRENT FILING DATE: 1999-06-08
;  EARLIER APPLICATION NUMBER: US 60/088,801
;  EARLIER FILING DATE: 1998-06-10
;  NUMBER OF SEQ ID NOS: 850
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 568
;   LENGTH: 616
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(616)
;   OTHER INFORMATION: n = A,T,C or G
US-09-328-111-568

  Query Match             1.8%;  Score 37.2;  DB 4;  Length 616;
  Best Local Similarity   58.3%;  Pred. No. 0.18;
  Matches   63;  Conservative    0;  Mismatches   45;  Indels    0;  Gaps    0;

Qy    1106 AGAAAAACCTTGACTTGCTTCTTCAAACAAAAAACAGTTCTAAAGTGCTTTTCAGAAAGA 1165

Db     290 AGACAGATCTGGTTTACCATCTTTTAACAAAAGGCAAATGTCTTCTCTTCTTCAGAAAGA 349

Qy    1166 GTGAAAAATACATGAAACTTTTTCTGTGGCTGCTTGTTGGTGTGGGAT 1213

Db     350 GTCATTAACACTAAAATTCTTTTCTTNNGAAGTTTCTTCTTTTCCGAT 397


RESULT 6
US-08-232-463-14
; Sequence 14, Application US/08232463
; Patent No. 5670367
;  GENERAL INFORMATION:
;    APPLICANT:  DORNER, F.
;    APPLICANT:  SCHEIFLINGER, F.
;    APPLICANT:  FALKNER, F. G.
;    TITLE OF INVENTION:  RECOMBINANT FOWLPOX VIRUS
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  1800 Diagonal Road, Suite 500
;      CITY:  Alexandria
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22313-0299
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/232,463
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/935,313
;      FILING DATE:
;      APPLICATION NUMBER:  EP 91 114 300.6
;      FILING DATE:  26-AUG-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BENT, Stephen A.
;      REGISTRATION NUMBER:  29,768
;      REFERENCE/DOCKET NUMBER:  30472/114 IMMU
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)836-9300
;      TELEFAX:  (703)683-4109
;      TELEX:  899149
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  7218 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      CLONE:  pTZgpt-F1s
US-08-232-463-14

  Query Match             1.8%;  Score 37.2;  DB 1;  Length 7218;
  Best Local Similarity   4.6%;  Pred. No. 0.68;
  Matches   18;  Conservative  203;  Mismatches  171;  Indels    0;  Gaps    0;

Qy     583 TGGCTGCTGGCAGGGGCTGCTTTTGGTAGCCTTGTGTTCCTCACCCACTGGGTTTTTGGA 642

Db    1040 TGGCTGCAGGTCGAGGGAGCTTGCGATYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1099

Qy     643 GAAGTCTCTCTTGTTTCCAGATGGGCAGTGAGTGGGCATCCACATCCAGGGCCAGATCCT 702

Db    1100 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1159

Qy     703 AACCCATTTGGAGGTGCAGTACTGCTGTGCTTGGCAAGTGGATTGATGCTTCCATCTTGT 762

Db    1160 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1219
```

```
Qy     763 TTGTGGTTTCGTGGTACTGGTTTGATCTGGTGGGTTACAGGAACAGCTTCAGCTGCGGGG 822
           :: :  :::: :  : ::  :::  :::  :   :: :     :  ::::  :: :
Db    1220 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1279

Qy     823 CTCCTTTACCTGCACACATGGGCAGCTGCTGTGTCTGGCTGTGTCTTCGCCATCTTTACT 882
           :::::::: ::: : : : :   :  :: :: : :::  :: : ::::: :: ::::: ::
Db    1280 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1339

Qy     883 GCATCCATGTGGCCCCAAACACTTGGACACCTTATTAACTCAGGGACAAACCCTGGGAAA 942
           : ::: : :  ::::   : :::   : :::: ::  :::     :   ::::
Db    1340 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1399

Qy     943 ACCATGACCATTGCCATGATATTTTATCTTCT 974
           :: :  :: :: :: :  : :::: ::::::
Db    1400 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1431


RESULT 7
US-08-359-705B-7/c
; Sequence 7, Application US/08359705B
; Patent No. 5844092
; GENERAL INFORMATION:
;    APPLICANT:  Presta, Leonard G.
;    APPLICANT:  Shelton, David L.
;    APPLICANT:  Urfer, Roman
;    TITLE OF INVENTION:  Human trk Receptors and Neurotrophic Factor Inhibitors
;    NUMBER OF SEQUENCES:  41
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  1 DNA Way
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 inch, 1.44 Mb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WinPatin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/359,705B
;      FILING DATE:  20-Dec-1994
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/286846
;      FILING DATE:  08/10/94
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/215139
;      FILING DATE:  03/18/94
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Torchia, PhD., Timothy E.
;      REGISTRATION NUMBER:  36,700
;      REFERENCE/DOCKET NUMBER:  P0873P2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650/225-8674
;      TELEFAX:  650/952-9881
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1858 base pairs
;      TYPE:  Nucleic Acid
;      STRANDEDNESS:  Single
;      TOPOLOGY:  Linear
US-08-359-705B-7

  Query Match             1.7%;  Score 34.8;  DB 2;  Length 1858;
  Best Local Similarity   58.8%;  Pred. No. 1.7;
  Matches   60;  Conservative    0;  Mismatches   42;  Indels    0;  Gaps    0;

Qy     580 AACTGGCTGCTGGCAGGGGCTGCTTTTGGTAGCCTTGTGTTCCTCACCCACTGGGTTTTT 639
           ||  |||| || |  |  |  ||  ||| | | || ||   |    |||| |||||||||
Db    1215 AAAGGGCTCCTTGAGGAAGTGGCCATTGATGGTCTGGTTGGCTGTGCCCAGTGGGTTTTT 1156

Qy     640 GGAGAAGTCTCTCTTGTTTCCAGATGGGCAGTGAGTGGGCAT 681
           ||  | |    | | ||| |||     | |||| |||||| |
Db    1155 GGCAATGAGGGTATAGTTGCCATTGTTGTAGTGGGTGGGCTT 1114


RESULT 8
US-08-286-846A-7/c
; Sequence 7, Application US/08286846A
; Patent No. 5877016
; GENERAL INFORMATION:
;    APPLICANT:  Presta, Leonard G.
;    APPLICANT:  Shelton, David L.
;    APPLICANT:  Urfer, Roman
;    TITLE OF INVENTION:  Human trk Receptors and Neurotrophic Factor Inhibitors
;    NUMBER OF SEQUENCES:  41
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  460 Point San Bruno Blvd
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 inch, 1.44 Mb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WinPatin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/286,846A
;      FILING DATE:  05-Aug-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Torchia, PhD., Timothy E.
;      REGISTRATION NUMBER:  36,700
;      REFERENCE/DOCKET NUMBER:  P0873P1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415/225-8674
;      TELEFAX:  415/952-9881
;      TELEX:  910/371-7168
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1858 base pairs
;      TYPE:  Nucleic Acid
;      STRANDEDNESS:  Single
;      TOPOLOGY:  Linear
US-08-286-846A-7

  Query Match             1.7%;  Score 34.8;  DB 2;  Length 1858;
  Best Local Similarity   58.8%;  Pred. No. 1.7;
  Matches   60;  Conservative    0;  Mismatches   42;  Indels    0;  Gaps    0;

Qy     580 AACTGGCTGCTGGCAGGGGCTGCTTTTGGTAGCCTTGTGTTCCTCACCCACTGGGTTTTT 639
           ||  |||| || |  |  |  ||  ||| | | || ||   |    |||| |||||||||
Db    1215 AAAGGGCTCCTTGAGGAAGTGGCCATTGATGGTCTGGTTGGCTGTGCCCAGTGGGTTTTT 1156

Qy     640 GGAGAAGTCTCTCTTGTTTCCAGATGGGCAGTGAGTGGGCAT 681
           ||  | |    | | ||| |||     | |||| |||||| |
Db    1155 GGCAATGAGGGTATAGTTGCCATTGTTGTAGTGGGTGGGCTT 1114


RESULT 9
US-08-457-880A-7/c
; Sequence 7, Application US/08457880A
; Patent No. 5910574
; GENERAL INFORMATION:
;    APPLICANT:  Leonard G. Presta
;    APPLICANT:  David L. Shelton
;    APPLICANT:  Roman Urfer
;    TITLE OF INVENTION:  HUMAN trk RECEPTORS AND NEUROTROPHIC FACTOR
;    TITLE OF INVENTION:  INHIBITORS
;    NUMBER OF SEQUENCES:  41
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  1 DNA Way
```

```
;       CITY:  South San Francisco
;       STATE:  California
;       COUNTRY:  USA
;       ZIP:  94080
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  3.5 inch, 1.44 Mb floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  WinPatin (Genentech)
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/457,880A
;       FILING DATE:
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/444,622
;       FILING DATE:  19-May-1995
;       APPLICATION NUMBER:  08/286846
;       FILING DATE:
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Torchia, PhD., Timothy E.
;       REGISTRATION NUMBER:  36,700
;       REFERENCE/DOCKET NUMBER:  P0873P1C3
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  650/225-8674
;       TELEFAX:  650/952-9881
;   INFORMATION FOR SEQ ID NO:  7:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1858 base pairs
;       TYPE:  Nucleic Acid
;       STRANDEDNESS:  Single
;       TOPOLOGY:  Linear
US-08-457-880A-7

  Query Match             1.7%;  Score 34.8;  DB 2;  Length 1858;
  Best Local Similarity   58.8%;  Pred. No. 1.7;
  Matches   60;  Conservative    0;  Mismatches   42;  Indels    0;  Gaps    0;

Qy     580 AACTGGCTGCTGGCAGGGGCTGCTTTTGGTAGCCTTGTGTTCCTCACCCACTGGGTTTTT 639
           ||  |||| || |  |  |  ||  ||| | | || ||   |    |||| |||||||||
Db    1215 AAAGGGCTCCTTGAGGAAGTGGCCATTGATGGTCTGGTTGGCTGTGCCCAGTGGGTTTTT 1156

Qy     640 GGAGAAGTCTCTCTTGTTTCCAGATGGGCAGTGAGTGGGCAT 681
           ||  | |    | | ||| |||     | |||| |||||| |
Db    1155 GGCAATGAGGGTATAGTTGCCATTGTTGTAGTGGGTGGGCTT 1114


RESULT 10
US-08-444-622A-7/c
; Sequence 7, Application US/08444622A
; Patent No. 6025166
;  GENERAL INFORMATION:
;    APPLICANT:  Leonard G. Presta
;    APPLICANT:  David L. Shelton
;    APPLICANT:  Roman Urfer
;    TITLE OF INVENTION:  HUMAN trk RECEPTORS AND NEUROTROPHIC FACTOR INHIBITORS
;    NUMBER OF SEQUENCES:  41
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  1 DNA Way
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 inch, 1.44 Mb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WinPatin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/444,622A
;      FILING DATE:  19-May-1995
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/286846
;      FILING DATE:  5
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Torchia, PhD., Timothy E.
;      REGISTRATION NUMBER:  36,700
;      REFERENCE/DOCKET NUMBER:  P0873P1C3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650/225-8674
;      TELEFAX:  650/952-9881
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1858 base pairs
;      TYPE:  Nucleic Acid
;      STRANDEDNESS:  Single
;      TOPOLOGY:  Linear
US-08-444-622A-7

  Query Match             1.7%;  Score 34.8;  DB 3;  Length 1858;
  Best Local Similarity   58.8%;  Pred. No. 1.7;
  Matches   60;  Conservative    0;  Mismatches   42;  Indels    0;  Gaps    0;

Qy     580 AACTGGCTGCTGGCAGGGGCTGCTTTTGGTAGCCTTGTGTTCCTCACCCACTGGGTTTTT 639
           ||  |||| || |  |  |  ||  ||| | | || ||   |    |||| |||||||||
Db    1215 AAAGGGCTCCTTGAGGAAGTGGCCATTGATGGTCTGGTTGGCTGTGCCCAGTGGGTTTTT 1156

Qy     640 GGAGAAGTCTCTCTTGTTTCCAGATGGGCAGTGAGTGGGCAT 681
           ||  | |    | | ||| |||     | |||| |||||| |
Db    1155 GGCAATGAGGGTATAGTTGCCATTGTTGTAGTGGGTGGGCTT 1114


RESULT 11
US-08-942-562-7/c
; Sequence 7, Application US/08942562
; Patent No. 6027927
;  GENERAL INFORMATION:
;    APPLICANT:  Presta, Leonard G.
;    APPLICANT:  Shelton, David L.
;    APPLICANT:  Urfer, Roman
;    TITLE OF INVENTION:  Human trk Receptors and Neurotrophic
;    TITLE OF INVENTION:  Factor Inhibitors
;    NUMBER OF SEQUENCES:  37
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  460 Point San Bruno Blvd
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 inch, 1.44 Mb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WinPatin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/942,562
;      FILING DATE:  01-OCT-1997
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/444,597
;      FILING DATE:  19-May-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Torchia, PhD., Timothy E.
;      REGISTRATION NUMBER:  36,700
;      REFERENCE/DOCKET NUMBER:  P0873P1C2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415/225-8674
;      TELEFAX:  415/952-9881
;      TELEX:  910/371-7168
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1858 base pairs
;      TYPE:  Nucleic Acid
;      STRANDEDNESS:  Single
```

```
;      TOPOLOGY:  Linear
US-08-942-562-7

  Query Match             1.7%;  Score 34.8;  DB 3;  Length 1858;
  Best Local Similarity   58.8%;  Pred. No. 1.7;
  Matches   60;  Conservative    0;  Mismatches   42;  Indels    0;  Gaps    0;

Qy     580 AACTGGCTGCTGGCAGGGGCTGCTTTTGGTAGCCTTGTGTTCCTCACCCACTGGGTTTTT 639
           ||  |||| || |  |  |  ||  ||| | | || ||   |    |||| |||||||||
Db    1215 AAAGGGCTCCTTGAGGAAGTGGCCATTGATGGTCTGGTTGGCTGTGCCCAGTGGGTTTTT 1156

Qy     640 GGAGAAGTCTCTCTTGTTTCCAGATGGGCAGTGAGTGGGCAT 681
           ||  | |    | | ||| |||     | |||| |||||| |
Db    1155 GGCAATGAGGGTATAGTTGCCATTGTTGTAGTGGGTGGGCTT 1114


RESULT 12
US-09-156-923-7/c
; Sequence 7, Application US/09156923
; Patent No. 6153189
;  GENERAL INFORMATION:
;    APPLICANT:  Presta, Leonard G.
;    APPLICANT:  Shelton, David L.
;    APPLICANT:  Urfer, Roman
;    TITLE OF INVENTION:  Human trk Receptors and Neurotrophic Factor Inhibitors
;    NUMBER OF SEQUENCES:  41
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Knobbe, Martens, Olson & Bear
;      STREET:  620 Newport Center Drive 16th Floor
;      CITY:  Newport Beach
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  92660
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 inch, 1.44 Mb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WinPatin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/156,923
;      FILING DATE:  18-SEP-1998
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/359,705
;      FILING DATE:  20-DEC-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/286846
;      FILING DATE:  10-AUG-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/215139
;      FILING DATE:  18-MAR-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Dreger, Ginger
;      REGISTRATION NUMBER:  33,055
;      REFERENCE/DOCKET NUMBER:  GENENT.33CP2C1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  949/760-0404
;      TELEFAX:  949/760-9502
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1858 base pairs
;      TYPE:  Nucleic Acid
;      STRANDEDNESS:  Single
;      TOPOLOGY:  Linear
US-09-156-923-7

  Query Match             1.7%;  Score 34.8;  DB 3;  Length 1858;
  Best Local Similarity   58.8%;  Pred. No. 1.7;
  Matches   60;  Conservative    0;  Mismatches   42;  Indels    0;  Gaps    0;

Qy     580 AACTGGCTGCTGGCAGGGGCTGCTTTTGGTAGCCTTGTGTTCCTCACCCACTGGGTTTTT 639
           ||  |||| || |  |  |  ||  ||| | | || ||   |    |||| |||||||||
Db    1215 AAAGGGCTCCTTGAGGAAGTGGCCATTGATGGTCTGGTTGGCTGTGCCCAGTGGGTTTTT 1156

Qy     640 GGAGAAGTCTCTCTTGTTTCCAGATGGGCAGTGAGTGGGCAT 681
           ||  | |    | | ||| |||     | |||| |||||| |
Db    1155 GGCAATGAGGGTATAGTTGCCATTGTTGTAGTGGGTGGGCTT 1114


RESULT 13
US-08-359-705B-5/c
; Sequence 5, Application US/08359705B
; Patent No. 5844092
;  GENERAL INFORMATION:
;    APPLICANT:  Presta, Leonard G.
;    APPLICANT:  Shelton, David L.
;    APPLICANT:  Urfer, Roman
;    TITLE OF INVENTION:  Human trk Receptors and Neurotrophic Factor Inhibitors
;    NUMBER OF SEQUENCES:  41
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  1 DNA Way
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 inch, 1.44 Mb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WinPatin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/359,705B
;      FILING DATE:  20-Dec-1994
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/286846
;      FILING DATE:  08/10/94
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/215139
;      FILING DATE:  03/18/94
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Torchia, PhD., Timothy E.
;      REGISTRATION NUMBER:  36,700
;      REFERENCE/DOCKET NUMBER:  P0873P2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650/225-8674
;      TELEFAX:  650/952-9881
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2715 base pairs
;      TYPE:  Nucleic Acid
;      STRANDEDNESS:  Single
;      TOPOLOGY:  Linear
US-08-359-705B-5

  Query Match             1.7%;  Score 34.8;  DB 2;  Length 2715;
  Best Local Similarity   58.8%;  Pred. No. 2.1;
  Matches   60;  Conservative    0;  Mismatches   42;  Indels    0;  Gaps    0;

Qy     580 AACTGGCTGCTGGCAGGGGCTGCTTTTGGTAGCCTTGTGTTCCTCACCCACTGGGTTTTT 639
           ||  |||| || |  |  |  ||  ||| | | || ||   |    |||| |||||||||
Db    1215 AAAGGGCTCCTTGAGGAAGTGGCCATTGATGGTCTGGTTGGCTGTGCCCAGTGGGTTTTT 1156

Qy     640 GGAGAAGTCTCTCTTGTTTCCAGATGGGCAGTGAGTGGGCAT 681
           ||  | |    | | ||| |||     | |||| |||||| |
Db    1155 GGCAATGAGGGTATAGTTGCCATTGTTGTAGTGGGTGGGCTT 1114


RESULT 14
US-08-286-846A-5/c
; Sequence 5, Application US/08286846A
; Patent No. 5877016
;  GENERAL INFORMATION:
;    APPLICANT:  Presta, Leonard G.
;    APPLICANT:  Shelton, David L.
```

```
;    APPLICANT:  Urfer, Roman
;    TITLE OF INVENTION:  Human trk Receptors and Neurotrophic Factor Inhibitors
;    NUMBER OF SEQUENCES:  41
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  460 Point San Bruno Blvd
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 inch, 1.44 Mb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WinPatin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/286,846A
;      FILING DATE:  05-Aug-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Torchia, PhD., Timothy E.
;      REGISTRATION NUMBER:  36,700
;      REFERENCE/DOCKET NUMBER:  P0873P1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415/225-8674
;      TELEFAX:  415/952-9881
;      TELEX:  910/371-7168
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2715 base pairs
;      TYPE:  Nucleic Acid
;      STRANDEDNESS:  Single
;      TOPOLOGY:  Linear
US-08-286-846A-5

  Query Match             1.7%;  Score 34.8;  DB 2;  Length 2715;
  Best Local Similarity   58.8%;  Pred. No. 2.1;
  Matches   60;  Conservative    0;  Mismatches   42;  Indels    0;  Gaps    0;

Qy     580 AACTGGCTGCTGGCAGGGGCTGCTTTTGGTAGCCTTGTGTTCCTCACCCACTGGGTTTTT 639
           ||  |||| || |  |  |  ||  ||| | | || ||   |    |||| |||||||||
Db    1215 AAAGGGCTCCTTGAGGAAGTGGCCATTGATGGTCTGGTTGGCTGTGCCCAGTGGGTTTTT 1156

Qy     640 GGAGAAGTCTCTCTTGTTTCCAGATGGGCAGTGAGTGGGCAT 681
           ||  | |    | | ||| |||     | |||| |||||| |
Db    1155 GGCAATGAGGGTATAGTTGCCATTGTTGTAGTGGGTGGGCTT 1114


RESULT 15
US-08-457-880A-5/c
; Sequence 5, Application US/08457880A
; Patent No. 5910574
;  GENERAL INFORMATION:
;    APPLICANT:  Leonard G. Presta
;    APPLICANT:  David L. Shelton
;    APPLICANT:  Roman Urfer
;    TITLE OF INVENTION:  HUMAN trk RECEPTORS AND NEUROTROPHIC FACTOR
;    TITLE OF INVENTION:  INHIBITORS
;    NUMBER OF SEQUENCES:  41
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  1 DNA Way
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 inch, 1.44 Mb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WinPatin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/457,880A
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/444,622
;      FILING DATE:  19-May-1995
;      APPLICATION NUMBER:  08/286846
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Torchia, PhD., Timothy E.
;      REGISTRATION NUMBER:  36,700
;      REFERENCE/DOCKET NUMBER:  P0873P1C3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650/225-8674
;      TELEFAX:  650/952-9881
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2715 base pairs
;      TYPE:  Nucleic Acid
;      STRANDEDNESS:  Single
;      TOPOLOGY:  Linear
US-08-457-880A-5

  Query Match             1.7%;  Score 34.8;  DB 2;  Length 2715;
  Best Local Similarity   58.8%;  Pred. No. 2.1;
  Matches   60;  Conservative    0;  Mismatches   42;  Indels    0;  Gaps    0;

Qy     580 AACTGGCTGCTGGCAGGGGCTGCTTTTGGTAGCCTTGTGTTCCTCACCCACTGGGTTTTT 639
           ||  |||| || |  |  |  ||  ||| | | || ||   |    |||| |||||||||
Db    1215 AAAGGGCTCCTTGAGGAAGTGGCCATTGATGGTCTGGTTGGCTGTGCCCAGTGGGTTTTT 1156

Qy     640 GGAGAAGTCTCTCTTGTTTCCAGATGGGCAGTGAGTGGGCAT 681
           ||  | |    | | ||| |||     | |||| |||||| |
Db    1155 GGCAATGAGGGTATAGTTGCCATTGTTGTAGTGGGTGGGCTT 1114


Search completed: October 21, 2002, 21:56:50
Job time : 88 secs
```

GenCore version 5.1.3

OM nucleic - nucleic search, using sw model

Run on: October 21, 2002, 18:11:18 ; Search time 2477 Seconds
(without alignments)
11459.062 Million cell updates/sec

Title: US-09-809-638-1
Perfect score: 2103
Sequence: 1 atgacctcgctgtggagaga.....ccaaatacttttttatgaaac 2103

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	736.2	35.0	1010	10	BG696417	BG696417 602659516
	2	524.8	25.0	872	10	BG247890	BG247890 602359805
	3	511	24.3	899	10	BI687517	BI687517 603315072
	4	492.4	23.4	748	10	BI854920	BI854920 603381910
	5	476.2	22.6	755	10	BI105048	BI105048 602892888
	6	456	21.7	585	10	BI343511	BI343511 371695 MA
	7	453.2	21.6	678	10	BI651562	BI651562 603298034
	8	445	21.2	596	10	BI343877	BI343877 372176 MA
	9	441.6	21.0	966	10	BG247983	BG247983 602359913
	10	436.6	20.8	597	9	AA710301	AA710301 vt53b05.r
c	11	418	19.9	631	9	AW235545	AW235545 xn19e05.x
	12	411.8	19.6	528	9	BE180939	BE180939 PM3-HT062
	13	409.4	19.5	974	10	BI853011	BI853011 603379834
c	14	395.8	18.8	617	9	AI962905	AI962905 wt24f04.x
	15	354.2	16.8	656	9	BB628092	BB628092 BB628092
	16	353.8	16.8	936	10	BF162283	BF162283 601770979
c	17	346.4	16.5	534	9	AW572339	AW572339 xq16f09.x
	18	337.4	16.0	352	9	AA369925	AA369925 EST81558
	19	332.4	15.8	378	10	BI001197	BI001197 PM0-HN007
	20	330.6	15.7	672	9	BB118197	BB118197 BB118197
c	21	324	15.4	510	9	AI377551	AI377551 tc15h10.x
c	22	322.4	15.3	520	9	AI812045	AI812045 tw46e03.x
	23	298.2	14.2	1079	10	BF164016	BF164016 601772165
c	24	286.4	13.6	485	9	AI373245	AI373245 qz48b10.x
c	25	243	11.6	439	9	AI717137	AI717137 UI-R-Y0-a
	26	231.4	11.0	445	10	BE375574	BE375574 601225462
c	27	230.8	11.0	408	10	BE328934	BE328934 hq32d07.x
c	28	223	10.6	290	9	AI868470	AI868470 tc51d12.x
	29	218.8	10.4	437	10	BG193159	BG193159 RST12284
c	30	202.2	9.6	394	9	AI826824	AI826824 wk56f12.x
	31	185.4	8.8	313	10	BG797173	BG797173 EST-01 Mo
c	32	180	8.6	495	9	AA863019	AA863019 og99f04.s
	33	176.6	8.4	425	9	BB846675	BB846675 BB846675
	34	175.4	8.3	428	9	BB845043	BB845043 BB845043
c	35	164.6	7.8	318	9	AA892501	AA892501 EST196304
c	36	140.4	6.7	370	9	AA523805	AA523805 ni68h11.s
c	37	139.6	6.6	998	12	CNS078KT	AL434115 T3 end of
	38	139.6	6.6	1115	12	CNS06PTX	AL409819 T7 end of
c	39	134.4	6.4	1056	12	CNS06MFW	AL405426 T3 end of
c	40	121.4	5.8	512	9	AW972717	AW972717 EST384809
c	41	117.8	5.6	1063	12	CNS06M3K	AL404982 T3 end of
c	42	112	5.3	305	9	AA195718	AA195718 zr33d07.s
	43	110.6	5.3	174	10	BF543325	BF543325 UI-R-Y0-a
	44	108.6	5.2	482	9	BB744625	BB744625 BB744625
	45	96.4	4.6	492	9	BB741693	BB741693 BB741693

ALIGNMENTS

RESULT 1
BG696417
LOCUS BG696417 1010 bp mRNA linear EST 07-MAY-2001
DEFINITION 602659516F1 NCI_CGAP_Skn3 Homo sapiens cDNA clone IMAGE:4802907 5', mRNA sequence.
ACCESSION BG696417
VERSION BG696417.1 GI:13961539
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1010)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: James Cleaver, M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM10697 row: o column: 04
High quality sequence stop: 803.
FEATURES Location/Qualifiers
source 1. .1010
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4802907"
/clone_lib="NCI_CGAP_Skn3"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: skin; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.5kb. Library constructed by Life Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT 318 a 208 c 230 g 254 t
ORIGIN

```
  Query Match             35.0%;  Score 736.2;  DB 10;  Length 1010;
  Best Local Similarity   97.7%;  Pred. No. 2.6e-195;
  Matches  800;  Conservative    0;  Mismatches   13;  Indels    6;  Gaps    5;

Qy   1289 CTGCCATCTGGCCTTTCAGGTTTGGATATGACAATGAAGGGTGGTCTAGTCTAGAAAGAT 1348
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 CTGCCATCTGGCCTTTCAGGTTTGGATATGACAATGAAGGGTGGTCTAGTCTAGAAAGAT 60

Qy   1349 CAGCTCACCT-GCTCAATGAAACAGGTGCAGATTTCATAACAATTTTGGAGAGTGATGCT 1407
          |||||||||| |||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 CAGCTCACCTCGCTCAATGAAACAGGTGCAGATTTCATAACAATTTTGGAGAGTGATGCT 120

Qy   1408 TCTAAGCCCTATATGGGGAACAATGACT-TAACCATGTGGCTAGGGGAAAAGTTGGGTTT 1466
          |||||||||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db    121 TCTAAGCCCTATATGGGGAACAATGACTCTAACCATGTGGCTAGGGGAAAAGTTGGGTTT 180

Qy   1467 CTATACAGACTTTGGTCCAAGCACAAGGTATCACACTTGGGGGATTATGGCTTTGTCAAG 1526
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 CTATACAGACTTTGGTCCAAGCACAAGGTATCACACTTGGGGGATTATGGCTTTGTCAAG 240

Qy   1527 ATACCCAATTGTGAAATCTGAGCATCACCTTCTTCCGTCACCAGAGGGCGAGATCGCACC 1586
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 ATACCCAATTGTGAAATCTGAGCATCACCTTCTTCCGTCACCAGAGGGCGAGATCGCACC 300

Qy   1587 AGCCATCACATTGACCGTTAACATTTCGGGCAAGCTGGTGGATTTTGTCGTGACACACTT 1646
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 AGCCATCACATTGACCGTTAACATTTCGGGCAAGCTGGTGGATTTTGTCGTGACACACTT 360

Qy   1647 TGGGAACCACGAAGATGACCTCGACAGGAAACTGCAGGCTATTGCTGTTTCAAAACTACT 1706
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 TGGGAACCACGAAGATGACCTCGACAGGAAACTGCAGGCTATTGCTGTTTCAAAACTACT 420

Qy   1707 GAAAAGTAGCTCTAATCAAGTGATATTTCTGGGATATATCACTTCAGCACCTGGCTCCAG 1766
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 GAAAAGTAGCTCTAATCAAGTGATATTTCTGGGATATATCACTTCAGCACCTGGCTCCAG 480

Qy   1767 AGATTATCTACAGCTCACTGAACATGGCAATGTGAAGGATATCGACAGCACTGATCATGA 1826
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 AGATTATCTACAGCTCACTGAACATGGCAATGTGAAGGATATCGACAGCACTGATCATGA 540

Qy   1827 CAGATGGTGTGAATACATTATGTATCGAGGGCTGATCAGGTTGGGTTATGCAAGAATCTC 1886
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 CAGATGGTGTGAATACATTATGTATCGAGGGCTGATCAGGTTGGGTTATGCAAGAATCTC 600

Qy   1887 CCATGCTGAACTG-AGTGATTCAGAAATTCAGATGGCAAAA-TTTAGGATCCCTGATGAC 1944
          |||||||||||||| |||||||||||||||||||||||||| |||||||||||||||||||
Db    601 CCATGCTGAACTGCAGTGATTCAGAAATTCAGATGGCAAAATTTTAGGATCCCTGATGAC 660

Qy   1945 CCCACTAATTATAGAGACAACCAGAAAGTGGTCATAGACCACAGAGAAGTTTCTGAGAAA 2004
          |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db    661 CCCACTAATTATAGAGACAAGCAGAAAGTGGTCATAGACCACAGAGAAGTTTCTGAGAAC 720

Qy   2005 ATTCATTTTAATCCCAGATTTGGATCCTACAAAGAAGGACACAATTATGAAAACAACCAT 2064
          ||||||||  |  ||||||| ||||||||||||||||||||||||||||||||||||||
Db    721 ATTCATTTGTAATCCAGATTGGGATCCTACAAAGAAGGACACAATTATGAAAACAAC--A 778

Qy   2065 CATTTTCATATGAATACTCCCAAATACTTTTTATGAAAC 2103
           | ||||||||||||||||||||| |||| |||| ||
Db    779 TAATTTCATATGAATACTCCCAAATAGTTTTGATGAGAC 817


RESULT 2
BG247890
LOCUS       BG247890                 872 bp    mRNA    linear   EST 13-FEB-2001
DEFINITION  602359805F1 NCI_CGAP_Mam1 Mus musculus cDNA clone IMAGE:4488289 5',
            mRNA sequence.
ACCESSION   BG247890
VERSION     BG247890.1  GI:12757705
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 872)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Gilbert Smith, Ph.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10334  row: j  column: 02
            High quality sequence stop: 730.
FEATURES             Location/Qualifiers
     source          1. .872
                     /organism="Mus musculus"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:4488289"
                     /clone_lib="NCI_CGAP_Mam1"
                     /tissue_type="tumor, biopsy sample"
                     /dev_stage="10 months, virgin"
                     /lab_host="DH10B"
                     /note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
                     Site_2: NotI; Cloned unidirectionally.  Primer: Oligo dT.
                     Library constructed by Life Technologies. Investigator
                     providing samples: Gilbert Smith, NIH"
BASE COUNT      229 a    190 c    225 g    228 t
ORIGIN

  Query Match             25.0%;  Score 524.8;  DB 10;  Length 872;
  Best Local Similarity   81.0%;  Pred. No. 5.5e-136;
  Matches  684;  Conservative    0;  Mismatches  152;  Indels    8;  Gaps    6;

Qy    913 CTTATTAACTCAGGGACAAACCCTGGGAAAACCATGACCATTGCCATGATATTTTATCTT 972
          ||| | |||||||||| |||| | ||  || ||||| ||| || ||||||  ||||| ||
Db      1 CTTGTGAACTCAGGGAAAAACTCCGGAGAAGCCATGGCCACTGGCATGATCCTTTATGTT 60

Qy    973 CTAGAAATATTTTTCTGTGCCTGGTGCACAGCTTTTAAGTTTGTCCCAGGAGGTGTCTAC 1032
          || ||| |||||||||||||||||||||||||||| |||||| |||| || ||||||||||*
Db     61 TTACAAACATTTTTCTGTGCCTGGTGCACAGCGTTTAAG-TTGTTCCCGGAGGTGTCTAC 119

Qy    1033 GCTAGAGAAAGATCAGATGTGCTTTTGGGGACAATGATGTTAATTATCGGGCTGAATATG 1092
           |||||||||||||| |||||||||||||||| || ||| ||||||| ||  ||| ||||
Db     120 GCTAGAGAAAGATCTGATGTGCTTTTGGGGACCATAATGGTAATTATTGGATTGAGTATG 179

Qy    1093 CTATTTGGTCCTAAGAAAAACCTTGACTTGCTTCTTCAAACAAAAAACAGTTCTAAAGTG 1152
           || ||||| |||||||||||||||||||| |||||||||||||||||||||| |||||  |
Db     180 CTGTTTGGACCTAAGAAAAACCTTGACTTTCTTCTTCAAACAAAAAACAGTCCTAAAACG 239

Qy    1153 CTTTTCAGAAAGAGTGAAAAATACATGAAACTTTTTCTGTGGCTGCTTGTTGGTGTGGGA 1212
           ||||| |||   |||||||||||||||||||  |||||||||| ||||||||||||||||
Db     240 CTTTTGAGATGCAGTGAAAAATACATGAAACTCATTCTGTGGCTGTTTGTTGGTGTGGGG 299

Qy    1213 TTGTTGGGATTAGGACTACGGCATAAAGCCTATGAGAGAAAACTGGGCAAAGTGGCACCA 1272
           ||  |||||||||| |||||||| | | |||| |||||  |  | |||  || |||||||
Db     300 TTACTGGGATTAGGGCTACGGCACAGAACCTACGAGAGGCAGTTAGGCCGAGGGGCACCA 359

Qy    1273 ACCAAAGAGGTCTCTGCTGCCATCTGGCCTTTCAGGTTTGGATATGACAATGAAGGGTGG 1332
           | | || |||||| |||||||||||||||||| ||||||||||||||||||||||||| |||
Db     360 GCGACAGTGGTCTCGGCTGCCATCTGGCCTTTTAGGTTTGGATATGACAATGAAGGCTGG 419

Qy    1333 TCTAGTCTAGAAAGATCAGCTCACCTGCTCAATGAAACAGGTGCAGATTTCATAACAATT 1392
           | | |||||| || || ||||| |||||||| || |||||||||||||| || || |||
Db     420 CCCAATCTAGAGAGGTCTGCTCAACTGCTCAAGGAGACAGGTGCAGATTTTATCACCATT 479

Qy    1393 TTGGAGAGTGATGCTTCTAAGCCCTATATGGGGAACAATGACTTAACCATGTGGCTAGGG 1452
           |||||||| |||||||| |||||||||| |||||||| |||||||| |||||||| |||
Db     480 TTGGAGAGCGATGCTTCCAAGCCCTATATAGGGAACAACGACTTAACAATGTGGCTGGGG 539
```

```
Qy   1453 GAAAAGTTGGGTTTCTATACAGACTTTGGTCCAAGCACAAGGTATCACACTTGGGGGAT- 1511
          || ||| |||||||||| |||||||||| |||||||| ||| ||||||| ||||  |
Db    540 GAGAAGCTGGGTTTCTACACAGACTTTGGGCCAAGCACCAGGGATCACACCTGGGGATTA 599

Qy   1512 TATGGCTTTGTCAAGATACCCAATTGTGAAATCTGAGCATCACCTTCTTCCGTC-ACCAG 1570
          |   ||| | ||   |  || |||||| | ||| || ||||||||||||||||  |||| 
Db    600 TGGTGCTGTCTCGGTACCCCGAATTGTCAGATCGGAACATCACCTTCTTCCGTCGGCCAG 659

Qy   1571 AGGGCGAGATCGCACCAGCCATCACATTGACCGTTAACATTTCGGGCAAGCTGGTGGATT 1630
          |||||||||||||||||||||| ||  |||| |||||| | ||   ||  |  || |  | 
Db    660 AGGGCGAGATCGCACCAGCCATAACCATGACAGTTAACGTCTCCAACAGACCTGTTGGAT 719

Qy   1631 TTGTCGTGACACACTTTGGGAACCACGAAGATGACCTCGACAGGAAACTGCAGGCTATTG 1690
          |||| |||||||   ||||||| |  |||||||||| |||||||| || |||||||| | 
Db    720 TTGTGGTGACACCACTTGGGAATCCTGAAGATGACCT-GACAGGAAGCTACAGGCTATGG 778

Qy   1691 CTGTTTCAAAACTACTGAAAAGTAGCTCTAATCAAGTGATATTTCTGGGATATATCACTT 1750
               ||| ||||| || |||| | |  |  ||| |||||||||| ||  |||| | | 
Db    779 --AGTTCCAAACTGCTAAAAACTCGTCC--ATCCAGTGATATTTCGGGTTTATACAATTA 834

Qy   1751 CAGC 1754
           | |
Db    835 AACC 838
```

RESULT 3
BI687517

LOCUS	BI687517 899 bp mRNA linear EST 18-SEP-2001
DEFINITION	603315072F1 NCI_CGAP_Mam6 Mus musculus cDNA clone IMAGE:5355137 5', mRNA sequence.
ACCESSION	BI687517
VERSION	BI687517.1 GI:15650145
KEYWORDS	EST.
SOURCE	house mouse.
ORGANISM	Mus musculus Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE	1 (bases 1 to 899)
AUTHORS	NIH-MGC http://mgc.nci.nih.gov/.
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL	Unpublished (1999)
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgapbs-r@mail.nih.gov Tissue Procurement: Jeffrey Green M.D. cDNA Library Preparation: Life Technologies, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: LLAM11902 row: p column: 18 High quality sequence stop: 786.

FEATURES	Location/Qualifiers
source	1. .899 /organism="Mus musculus" /strain="FVB/N" /db_xref="taxon:10090" /clone="IMAGE:5355137" /clone_lib="NCI_CGAP_Mam6" /sex="female, virgin" /tissue_type="infiltrating ductal carcinoma" /dev_stage="5 months" /lab_host="DH10B" /note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI; Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT. Library constructed by Life Technologies. Investigator providing samples: Jeffrey Green, M.D., NIH"

BASE COUNT 170 a 247 c 247 g 235 t
ORIGIN

Query Match 24.3%; Score 511; DB 10; Length 899;
Best Local Similarity 76.8%; Pred. No. 4.2e-132;
Matches 637; Conservative 0; Mismatches 190; Indels 2; Gaps 1;

```
Qy      1 ATGACCTCGCTGTGGAGAGAAATCCTCTTGGAGTCGCTGCTGGGATGTGTTTCTTGGTCT 60
          ||| |    |||||||||| |||   | ||||| |  ||||||| |  || || ||||||
Db     69 ATGCCAGGCCTGTGGAGAGCAATTGCCCTGGAGACCTTGCTGGGCTACGTCTCCTGGTCT 128

Qy     61 CTCTACCATGACCTGGGACCGATGATCTATTACTTTCCTTTGCAAACACTAGAACTCACT 120
          ||||||||  |  ||| | || |||||||||| |||||  ||||||| || || |||||
Db    129 CTCTACCACGGACTGAGCCCCATGATCTATTATTTTCCCCTGCAAACGCTGGAGCTCACA 188

Qy    121 GGGCTTGAAGGTTTTAGTATAGCATTTCTTTCTCCAATATTCCTAACAATTACTCCTTTC 180
          |||||  |||    ||| || |||| |||||  |||||  ||| || ||||||||  |  ||  ||
Db    189 GGGCTGGAATTCTTTTGTGTAGCCTTTCTCTCTCCGATACTCTTAACAATTCCCCCACTC 248

Qy    181 TGGAAATTGGTTAACAAGAAGTGGATGCTAACCCTGCTGAGGATAATCACTATTGGCAGC 240
          ||||||  |  ||  || ||||||||||   ||   | ||||||||||   |||||  | ||||||
Db    249 TGGAAACTCGTCAATAAGAAGTGGACACTCTCTCTGCTGAGGATCGTCACTGTAGGCAGC 308

Qy    241 ATAGCCTCCTTCCAGGCTCCAAATGCCAAACTTCGACTGATGGTTCTTGCGCTTGGGGTG 300
          |||||  |||||  ||||||||||||||||  ||||||||||||||  ||  ||  |||||  |||
Db    309 ATAGCTTCCTTTGAGGCTCCAAATGCCAAGCTTCGACTGATGGTCCTCGCCCTTGGTGTG 368

Qy    301 TCTTCCTCACTGATAGTGCAAGCTGTGACTTGGTGGTCGGGAAGTCATTTGCAAAGGTAC 360
          |||||  ||  |||||||||||  |||||||  ||||||||  ||||||   ||||||||||||
Db    369 TCTTCTTCCTTGATAGTGCAAACTGTGACCTGGTGGTCAGGAAGTGGATTGCAAAGGTAC 428

Qy    361 CTCAGAATTTGGGGATTCATTTTAGGACAGATTGTTCTTGTTGTTCTACGCATATGGTAT 420
          ||||  |||  ||||||||||||  |  |||||  |  |||||  ||||  ||  ||  ||||||||
Db    429 CTCAAAATCTGGGGATTCATTCTGGGACACGTCCTTCTTCTTGTCCTCCGAATATGGTAC 488

Qy    421 ACTTCACTAAACCCAATCTGGAGTTATCAGATGTCCAACAAAGTGATACTGACATTAAGT 480
          ||  |||  ||||||||||  |||||  ||||||||||||||||  ||||||  ||||||||  |||
Db    489 ACATCATTAAACCCAATTTGGAGCTATCAGATGTCCAACAGAGTGATTCTGACATTGAGT 548

Qy    481 GCCATAGCCACACTTGATCGTATTGGCACAGATGGTGACTGCAGTAAACCTGAAGAAAAG 540
          |||  ||||    |||||  ||  ||||||||||||||||||||  |||  ||  |||||  |  |||
Db    549 GCCGTAGCTGTGCTTGACCGAATTGGCACAGATGGTGACTACAGAAACCCTGAGGGGAAG 608

Qy    541 AAGACTGGTGAGGTAGCCACGGGGATGGCCTCTAGACCCAACTGGCTGCTGGCAGGGGCT 600
          |||  |  |||||||  ||||||||  |  |  |||||  |  |  |  ||||||  ||  ||||  ||
Db    609 AAGCCCCGTGAGGTTGCCACGGGCAGGACCTCTCTATCTAGTTGGCTGTTGCCAGGCGCC 668

Qy    601 GCTTTTGGTAGCCTTGTGTTCCTCACCCACTGGGTTTTTGGAGAAGTCTCTCTTGTTTCC 660
          ||  |||||  |||||  |||||||  ||  ||||||  |  ||  |||||  |||||  ||||||||
Db    669 GCCTTTGGGAGCCTCCTGTTCCTAACTCACTGGATCTTCGGAGAGGTCTCCATTGTTTCC 728

Qy    661 AGATGGGCAGTGAGTGGGCATCCACATCCAGGGCCAGATCCTAACCCATTTGGAGGTGCA 720
          ||||||||||  |||||  |||||  ||  ||||||||||||||||  ||  ||||||||||
Db    729 AGATGGGCAGTCAGTGGTCATCCCCACCCAGGGCCAGATCCTAATCCTTTTGGAGGTGCC 788

Qy    721 GTACTGCTGTGCTTGGCAAGTGGATTGATGCTTCC--ATCTTGTTTGTGGTTTCGTGGTA 778
          |  ||  |||  ||||  ||||||||  |  |||    |      |       ||||  |  | |
Db    789 GCTCTACTGGGCTTCTCAAGTGGACTTGTGCCAACGGGGTTCACCATGGGTTCCTGGATG 848

Qy    779 CTGGTTTGATCTGGTGGGTTACAGGAACAGCTTCAGCTGCGGGGCTCCT 827
          ||||  ||   ||||||||||  |  ||  ||  |||  |    |||   |||
Db    849 CTGGGTTAGCCTGGTGGGTTCCCGGGCCACTTTCGGGCCGGGGTTCCCT 897
```

RESULT 4
BI854920

LOCUS	BI854920 748 bp mRNA linear EST 10-OCT-2001
DEFINITION	603381910F1 NCI_CGAP_Mam6 Mus musculus cDNA clone IMAGE:5397050 5', mRNA sequence.
ACCESSION	BI854920
VERSION	BI854920.1 GI:15995667
KEYWORDS	EST.
SOURCE	house mouse.
ORGANISM	Mus musculus Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE	1 (bases 1 to 748)
AUTHORS	NIH-MGC http://mgc.nci.nih.gov/.
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)

```
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Jeffrey Green M.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM12012  row: c  column: 03
            High quality sequence stop: 746.
FEATURES             Location/Qualifiers
     source          1. .748
                     /organism="Mus musculus"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:5397050"
                     /clone_lib="NCI_CGAP_Mam6"
                     /sex="female, virgin"
                     /tissue_type="infiltrating ductal carcinoma"
                     /dev_stage="5 months"
                     /lab_host="DH10B"
                     /note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
                     Site_2: NotI; Cloned unidirectionally.  Primer: Oligo dT.
                     Library constructed by Life Technologies. Investigator
                     providing samples: Jeffrey Green, M.D., NIH"
BASE COUNT      198 a    162 c    207 g    181 t
ORIGIN

  Query Match           23.4%;  Score 492.4;  DB 10;  Length 748;
  Best Local Similarity   84.2%;  Pred. No. 6.5e-127;
  Matches  625;  Conservative    0;  Mismatches  111;  Indels    6;  Gaps    6;

Qy   1189 CTGTGGCTGCTTGTTGGTGTGGGATTGTTGGGATTAGGACTACGGCATAAAGCCTATGAG 1248
           ||||||||| |||||||||||||| ||  |||||||||| |||||||| | | |||| |||
Db      1 CTGTGGCTGTTTGTTGGTGTGGGGTTACTGGGATTAGGGCTACGGCACAGAACCTACGAG 60

Qy   1249 AGAAAACTGGGCAAAGTGGCACCAACCAAAGAGGTCTCTGCTGCCATCTGGCCTTTCAGG 1308
           ||  |  | |||  || |||||| | | || |||||| ||||||||||||||||| |||
Db     61 AGGCAGTTAGGCCGAGGGGCACCAGCGACAGTGGTCTCGGCTGCCATCTGGCCTTTTAGG 120

Qy   1309 TTTGGATATGACAATGAAGGGTGGTCTAGTCTAGAAAGATCAGCTCACCTGCTCAATGAA 1368
           |||||||||||||||||||| ||| | | |||||| || || ||||| |||||||| ||
Db    121 TTTGGATATGACAATGAAGGCTGGCCCAATCTAGAGAGGTCTGCTCAACTGCTCAAGGAG 180

Qy   1369 ACAGGTGCAGATTTCATAACAATTTTGGAGAGTGATGCTTCTAAGCCCTATATGGGGAAC 1428
           |||||||||||||| || || | ||||||||| |||||||| |||||||||| ||||||
Db    181 ACAGGTGCAGATTTTATCACCA-TTTGGAGAGCGATGCTTCCAAGCCCTATATAGGGAAC 239

Qy   1429 AATGACTTAACCATGTGGCTAGGGGAAAAGTTGGGTTTCTATACAGACTTTGGTCCAAGC 1488
           || |||||||| |||||||| ||||| ||| |||||||||| |||||||||| ||||||
Db    240 AACGACTTAACAATGTGGCTGGGGGAGAAGCTGGGTTTCTACACAGACTTTGGGCCAAGC 299

Qy   1489 ACAAGGTATCACACTTGGGGGATTATGGCTTTGTCAAGATACCCAATTGTGAAATCTGAG 1548
           || ||| ||||||| |||||||||||||    ||||  | ||||| | ||| | ||| ||
Db    300 ACCAGG-ATCACACCTGGGGGATTATGGTGCTGTCTCGGTACCCGA-TGTCAGATCGGAA 357

Qy   1549 CATCACCTTCTTCCGTCACCAGAGGGCGAGATCGCACCAGCCATCACATTGACCGTTAAC 1608
           |||||||||||||||| |||||||||||||||||||||||||| ||  |||| ||||||
Db    358 CATCACCTTCTTCCGTCGCCAGAGGGCGAGATCGCACCAGCCATAACCATGACAGTTAAC 417

Qy   1609 ATTTCGGGCAAGCTGGTGGATTTTGTCGTGACACACTTTGGGAACCACGAAGATGACCTC 1668
           | ||   ||  |||||||| ||||| ||||||||||||||||| || ||||||||||
Db    418 GTCTCCAACAGACTGGTGGA-TTTGTGGTGACACACTTTGGGAATCATGAAGATGACCTT 476

Qy   1669 GACAGGAAACTGCAGGCTATTGCTGTTTCAAAACTACTGAAAAGTAGCTCTAATCAAGTG 1728
           |||||||| || |||||||||| |||||||||| ||||||| | | || |||||||||
Db    477 GACAGGAAGCTACAGGCTATTGCAGTTTCAAAACTGCTGAAAAATTGTTCAAATCAAGTG 536

Qy   1729 ATATTTCTGGGATATATCACTTCAGCACCTGGCTCCAGAGA-TTATCTACAGCTCACTGA 1787
           |||||||||| |||||||||||||  ||||| |||||||| |||  |||| |||||  |
Db    537 ATATTTCTGGGTTATATCACTTCAGAGCCTGGTTCCAGAGATTTACATACAACTCACAAA 596

Qy   1788 ACATGGCAATGTGAAGGATATCGACA-GCACTGATCATGACAGATGGTGTGAATACATTA 1846
           ||||||||| |||||||||||| |||| || | |||   |||||||||||||| ||||| |
Db     597 ACATGGCAACGTGAAGGATATTGACACGCTCAGATGGAGACAGATGGTGTGAGTACATCA 656

Qy   1847 TGTATCGAGGGCTGATCAGGTTGGGTTATGCAAGAATCTCCCATGCTGAACTGAGTGATT 1906
           |||| || || |||||||||||||| |||||||||||||| |||| ||||||||||| |
Db     657 TGTACCGGGGTCTGATCAGGTTGGGCTATGCAAGAATCTCTCATGCAGAACTGAGTGACT 716

Qy   1907 CAGAAATTCAGATGGCAAAATT 1928
           | || |||||||||| |||||
Db     717 CTGAGATTCAGATGGCCAAATT 738


RESULT 5
BI105048
LOCUS       BI105048                 755 bp    mRNA    linear   EST 26-JUN-2001
DEFINITION  602892888F1 NCI_CGAP_Lu29 Mus musculus cDNA clone IMAGE:5037956 5',
            mRNA sequence.
ACCESSION   BI105048
VERSION     BI105048.1  GI:14555941
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 755)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Gilbert Smith, Ph.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11104  row: p  column: 21
            High quality sequence stop: 739.
FEATURES             Location/Qualifiers
     source          1. .755
                     /organism="Mus musculus"
                     /strain="CZECH II (feral)"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:5037956"
                     /clone_lib="NCI_CGAP_Lu29"
                     /tissue_type="spontaneous tumor, metastatic to mammary.
                     Stem cell origin."
                     /lab_host="DH10B"
                     /note="Organ: lung; Vector: pCMV-SPORT6; Site_1: SalI;
                     Site_2: NotI; Cloned unidirectionally.  Primer: Oligo dT.
                     Library constructed by Life Technologies. Investigator
                     providing samples: Gilbert Smith, NIH"
BASE COUNT      176 a    165 c    205 g    209 t
ORIGIN

  Query Match           22.6%;  Score 476.2;  DB 10;  Length 755;
  Best Local Similarity   81.1%;  Pred. No. 2.3e-122;
  Matches  614;  Conservative    0;  Mismatches  138;  Indels    5;  Gaps    5;

Qy    683 CACATCCAGGGCCAGATCCTAACCCATTTGGAGGTGCAGTACTGCTGTGCTTGGCAAGTG 742
          | |  ||  |||| |||||||| || |||||||||| || || ||| ||||  ||||||
Db      2 CCCCACCCAGGCCCGATCCTAATCCTTTTGGAGGTGCGGTTCTACTGGGCTTCTCAAGTG 61

Qy    743 GATTGATGCTTCCATCTTGTTTGTGGTTTCGTGGTACTGGTTTGATCTGGTGGGTTACAG 802
          || |||||||  |   ||  |  ||| ||| || | |||| ||   ||||||| | ||||
Db     62 GACTGATGCTATCGGGTTCATCATGGCTTCATGATGCTGGGTTAGCCTGGTGGATGACAG 121

Qy    803 GAACAGCTTCAGCTGCGGGGCTCCTTTACCTGCACACATGGGCAGCTGCTGTGTCTGGCT 862
          || ||||||| ||   ||| ||||| ||||||| |||||||||||||||||| ||||| |
Db    122 GAGCAGCTTCGGCCATGGGTCTCCTCTACCTGCGCACATGGGCAGCTGCTGTTTCTGGAT 181
```

```
Qy     863 GTGTCTTCGCCATCTTTACTGCATCCATGTGGCCCCAAACACTTGGACACCTTATTAACT 922
           ||||| | ||  ||||||| |  || |||||||| |||  |||||| || ||| | ||||
Db     182 GTGTCCTGGCAGTCTTTACGGGGTCTATGTGGCCTCAAGTACTTGGCCATCTTGTGAACT 241

Qy     923 CAGGGACAAACCCTGGGAAAACCATGACCATTGCCATGATATTTTATCTTCTAGAAATAT 982
           |||||| |||| | ||  || ||||| ||| || ||||||  ||||| || || ||| ||
Db     242 CAGGGAAAAACTCCGGAGAAGCCATGGCCACTGGCATGATCCTTTATGTTTTACAAACAT 301

Qy     983 TTTTCTGTGCCTGGTGCACAGCTTTTAAGTTTGTCCCAGGAGGTGTCTACGCTAGAGAAA 1042
           ||||||||||||||||||||| |||||| |||| || ||||||||||||||||||||||
Db     302 TTTTCTGTGCCTGGTGCACAGCGTTTAAG-TTGTTCCCGGAGGTGTCTACGCTAGAGAAA 360

Qy    1043 GATCAGATGTGCTTTTGGGGACAATGATGTTAATTATCGGGCTGAATATGCTATTTGGTC 1102
           |||| ||||||| ||||||||| || ||| ||||||| ||  ||| |||||| ||||| |
Db     361 GATCTGATGTGC-TTTGGGGACCATAATGGTAATTATTGGATTGAGTATGCTGTTTGGAC 419

Qy    1103 CTAAGAAAAACCTTGACTTGCTTCTTCAAACAAAAAACAGTTCTAAAGTGCTTTTCAGAA 1162
           |||||||||||||||||| |||||||||| ||| ||||||||| || | |||||
Db     420 CTAAGAAAAACCTTGACTTTCTTCTTCAAACAAAAAACAGTCCTAAAACGCTTTTGAGAT 479

Qy    1163 AGAGTGAAAAATACATGAAACTTTTTCTGTGGCTGCTTGTTGGTGTGGGATTGTTGGGAT 1222
             ||||||||||||||||||  |||||||||| ||| |||||||| || | |||||
Db     480 GCAGTGAAAAATACATGAAACTCATTCTGTGGCTGTTTG-TGGTGTGGGGTTATCGGGAT 538

Qy    1223 TAGGACTACGGCATAAAGCCTATGAGAGAAAACTGGGCAAAGTGGCACCAACCAAAGAGG 1282
           |||| |||||||| | | |||| |||||  |  | |||  || ||||||| | | || ||
Db     539 TAGGGCTACGGCACAGAACCTACGAGAGGCAGTTAGGCCGAGGGGCACCAGCGACAGTGG 598

Qy    1283 TCTCTGCTGCCATCTGGCCTTTCAGGTTTGGATATGACAATGAAGGGTGGTCTAGTCTAG 1342
           |||| |||||||||||||||| |||||||||||||||||| ||||| ||| | | |||||
Db     599 TCTCGGCTGCCATCTGGCCTTTTAGGTTTGGATATGACAACGAAGGCTGGCCCAATCTAG 658

Qy    1343 AAAGATCAGCTCACCTGCTCAATGAAACAGGTGCAGA-TTTCATAACAATTTTGGAGAGT 1401
           | || || | ||| |||||||| || |||||||||| ||| || || ||||||||||
Db     659 AGAGGTCTG-TCAACTGCTCAAGGAGACAGGTGCAGATTTTTATCACCATTTTGGAGAGC 717

Qy    1402 GATGCTTCTAAGCCCTATATGGGGAACAATGACTTAA 1438
           |||||||| |||||||||| |||||||| |||||||
Db     718 GATGCTTCCAAGCCCTATATAGGGAACAACGACTTAA 754
```

RESULT 6
BI343511
LOCUS BI343511 585 bp mRNA linear EST 30-JUL-2001
DEFINITION 371695 MARC 2PIG Sus scrofa cDNA 5', mRNA sequence.
ACCESSION BI343511
VERSION BI343511.1 GI:15036800
KEYWORDS EST.
SOURCE pig.
ORGANISM Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE 1 (bases 1 to 585)
AUTHORS Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Smith,T.P.L., Casas,E., Stone,R.T., Heaton,M.P., Grosse,W.M., Bennett,G.A., Laegreid,W.W. and Keele,J.W.
TITLE Design and use of two pooled tissue normalized cDNA libraries for EST discovery in swine
JOURNAL Unpublished (2000)
COMMENT Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and alt_trimmed with phred v0.980904.e. Vector identified by cross_match with the -minscore 18 and -minmatch 12 options.
PCR PRimers
FORWARD: AGGAAACAGCTATGACCAT
BACKWARD: GTTTTCCCAGTCACGACG
Plate: 116 row: N column: 24
Seq primer: ATTTAGGTGACACTATAG.
FEATURES Location/Qualifiers
source 1. .585
/organism="Sus scrofa"
/db_xref="taxon:9823"
/clone_lib="MARC 2PIG"
/tissue_type="pooled"
/lab_host="DH10B"
/note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI; Library made from pooled tissue from testis, ovary, endometrium, hypothalamus, pituitary, and placenta."
BASE COUNT 185 a 132 c 124 g 144 t
ORIGIN

Query Match 21.7%; Score 456; DB 10; Length 585;
Best Local Similarity 89.1%; Pred. No. 9.7e-117;
Matches 492; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

```
Qy    1550 ATCACCTTCTTCCGTCACCAGAGGGCGAGATCGCACCAGCCATCACATTGACCGTTAACA 1609
           ||||||||||||||||||||||||||||||||||| ||||||||||||||  |||| || | ||
Db       1 ATCACCTTCTTCCGTCACCAGAGGGCGAGATTGCACCAGCCATCACGCTGACTGTCACCA 60

Qy    1610 TTTCGGGCAAGCTGGTGGATTTTGTCGTGACACACTTTGGGAACCACGAAGATGACCTCG 1669
           ||||   || |||||||| |||||||||| || ||||||||  ||||||||||||||||
Db      61 TTTCCAACAGACTGGTGGACTTTGTCGTGACCCATTTTGGGAATGACGAAGATGACCTCG 120

Qy    1670 ACAGGAAACTGCAGGCTATTGCTGTTTCAAAACTACTGAAAAGTAGCTCTAATCAAGTGA 1729
           |||| ||||| || |||||||| || ||||||||||||||| ||||||| | |||||||
Db     121 ACAGAAAACTTCAAGCTATTGCCGTCTCAAAACTACTGAAAAATAGCTCTGAGCAAGTGA 180

Qy    1730 TATTTCTGGGATATATCACTTCAGCACCTGGCTCCAGAGATTATCTACAGCTCACTGAAC 1789
           |||||||||||| |||||||||  |||||| |||||||||||||| |||||||| |||||
Db     181 TATTTCTGGGATACATCACTTCAAAACCTGGTTCCAGAGATTATCGACAGCTCATTGAAC 240

Qy    1790 ATGGCAATGTGAAGGATATCGACAGCACTGATCATGACAGATGGTGTGAATACATTATGT 1849
           |||||||||||||||||||||||||||||||| || |  |||||||||||||||| |||||||
Db     241 ATGGCAATGTGAAGGATATCGACAGCACAGACCGAGACAGATGGTGTGAATATATTATGT 300

Qy    1850 ATCGAGGGCTGATCAGGTTGGGTTATGCAAGAATCTCCCATGCTGAACTGAGTGATTCAG 1909
           | |||||||||||||  |||||||||||||||||||| |||||||||||||| || |
Db     301 ACCGAGGGCTGATCAGACTGGGTTATGCAAGAATCTCCCGTGCTGAACTGAGTGACTCTG 360

Qy    1910 AAATTCAGATGGCAAAATTTAGGATCCCTGATGACCCCACTAATTATAGAGACAACCAGA 1969
           |||||||||||| |||||||||||||||||||||||  ||||||||| ||||||||||
Db     361 AAATTCAGATGGCCAAATTTAGGATCCCTGATGACCCCCTTAATTATAGGGACAACCAGA 420

Qy    1970 AAGTGGTCATAGACCACAGAGAAGTTTCTGAGAAAATTCATTTTAATCCCAGATTTGGAT 2029
           | ||||||| |||||||||  ||||| ||||||| |||||||| |||||||| ||||| |
Db     421 AGGTGGTCACAGACCACAGGAAAGTTCCTGAGAAGATTCATTTCAATCCCAGGTTTGGGT 480

Qy    2030 CCTACAAAGAAGGACACAATTATGAAAACAACCATCATTTTCATATGAATACTCCCAAAT 2089
           |||||||||||||||||||  ||| ||||| |||||||||| |||| ||||||||||
Db     481 CCTACAAAGAAGGACACAATGCTGAGAACAATCATCATTTTCACATGAGTACTCCCAAAT 540

Qy    2090 ACTTTTTATGAA 2101
           ||||||||||||
Db     541 ACTTTTTATGAA 552
```

RESULT 7
BI651562
LOCUS BI651562 678 bp mRNA linear EST 12-SEP-2001
DEFINITION 603298034F1 NIH_CGAP_Mam3 Mus musculus cDNA clone IMAGE:5338688 5', mRNA sequence.
ACCESSION BI651562
VERSION BI651562.1 GI:15565798
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 678)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.

```
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Lothar Hennighausen Ph.D., Chu-Xia Deng Ph.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11860  row: c  column: 09
            High quality sequence stop: 676.
FEATURES             Location/Qualifiers
     source          1. .678
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:5338688"
                     /clone_lib="NIH_CGAP_Mam3"
                     /tissue_type="tumor, gross tissue"
                     /lab_host="DH10B"
                     /note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                     Average insert 2 kb. Library constructed by Life
                     Technologies, catalog #12017-018. Investigators providing
                     samples: Lothar Hennighausen/Chu-Xia Deng, NIH Reference
                     for transgenic model: Xu et al., Nature Genetics 22, 37-43
                     (1999). Note: this is a NCI_CGAP Library."
BASE COUNT      139 a    156 c    191 g    192 t
ORIGIN

  Query Match             21.6%;  Score 453.2;  DB 10;  Length 678;
  Best Local Similarity   80.2%;  Pred. No. 6.3e-116;
  Matches  544;  Conservative    0;  Mismatches  133;  Indels    1;  Gaps    1;

Qy    437 TCTGGAGTTATCAGATGTCCAACAAAGTGATACTGACATTAAGTGCCATAGCCACACTTG 496
          | ||||| ||||||||||||||||| ||| || |||||||| |||||| ||||    ||||
Db      1 TTTGGAGCTATCAGATGTCCAACAGAGTAATTCTGACATTGAGTGCCGTAGCTGTGCTTG 60

Qy    497 ATCGTATTGGCACAGATGG-TGACTGCAGTAAACCTGAAGAAAAGAAGACTGGTGAGGTA 555
          | || |||||||||||||| ||||| ||| || ||||| |  ||||||  |  |||||||
Db     61 ACCGAATTGGCACAGATGGTTGACTACAGAAACCCTGAGGGGAAGAAGCCCCGTGAGGTT 120

Qy    556 GCCACGGGGATGGCCTCTAGACCCAACTGGCTGCTGGCAGGGGCTGCTTTTGGTAGCCTT 615
          |||||||| | | |||||  | | |  |||||| || |||| || || ||||| |||||
Db    121 GCCACGGGCAGGACCTCTCTATCTAGTTGGCTGTTGCCAGGCGCCGCCTTTGGGAGCCTC 180

Qy    616 GTGTTCCTCACCCACTGGGTTTTTGGAGAAGTCTCTCTTGTTTCCAGATGGGCAGTGAGT 675
          ||||||| || |||||| | || ||||| |||||  ||||||||||||||||||| |||
Db    181 CTGTTCCTAACTCACTGGATCTTCGGAGAGGTCTCCATTGTTTCCAGATGGGCAGTCAGT 240

Qy    676 GGGCATCCACATCCAGGGCCAGATCCTAACCCATTTGGAGGTGCAGTACTGCTGTGCTTG 735
          || ||||| || ||||||||||||||||| || |||||||||| || || ||| ||||
Db    241 GGTCATCCCCACCCAGGGCCAGATCCTAATCCTTTTGGAGGTGCGGTTCTACTGGGCTTC 300

Qy    736 GCAAGTGGATTGATGCTTCCATCTTGTTTGTGGTTTCGTGGTACTGGTTTGATCTGGTGG 795
          |||||||| |||||||  |   ||  |  ||| ||| || | |||| ||   |||||||
Db    301 TCAAGTGGACTGATGCTATCGGGTTCATCATGGCTTCATGATGCTGGGTTAGCCTGGTGG 360

Qy    796 GTTACAGGAACAGCTTCAGCTGCGGGGCTCCTTTACCTGCACACATGGGCAGCTGCTGTG 855
          | |||||| ||||||| ||   ||| ||||| ||||||| ||||||||||||||||||
Db    361 ATGACAGGAGCAGCTTCGGCCATGGGTCTCCTCTACCTGCGCACATGGGCAGCTGCTGTT 420

Qy    856 TCTGGCTGTGTCTTCGCCATCTTTACTGCATCCATGTGGCCCCAAACACTTGGACACCTT 915
          ||||| |||||| | ||  ||||||| |  || |||||||| |||  |||||| || |||
Db    421 TCTGGATGTGTCCTGGCAGTCTTTACGGGGTCTATGTGGCCTCAAGTACTTGGCCATCTT 480

Qy    916 ATTAACTCAGGGACAAACCCTGGGAAAACCATGACCATTGCCATGATATTTTATCTTCTA 975
          | |||||||||| |||| | ||  || ||||| ||| || ||||||  ||||| || ||
Db    481 GTGAACTCAGGGAAAAACTCCGGAGAAGCCATGGCCACTGGCATGATCCTTTATGTTTTA 540

Qy    976 GAAATATTTTTCTGTGCCTGGTGCACAGCTTTTAAGTTTGTCCCAGGAGGTGTCTACGCT 1035
          ||| ||||||||||||||||||||||||| |||||||||| || ||||||||||||||||
```

```
Db    541 CAAACATTTTTCTGTGCCTGGTGCACAGCGTTTAAGTTTGTTCCCGGAGGTGTCTACGCT 600

Qy   1036 AGAGAAAGATCAGATGTGCTTTTGGGGACAATGATGTTAATTATCGGGCTGAATATGCTA 1095
          |||||||||||| |||||||||||||||||| || ||| ||||||| ||  ||| ||||||
Db    601 AGAGAAAGATCTGATGTGCTTTTGGGGACCATAATGGTAATTATTGGATTGAGTATGCTG 660

Qy   1096 TTTGGTCCTAAGAAAAAC 1113
          ||||| ||||||||||||
Db    661 TTTGGACCTAAGAAAAAC 678


RESULT 8
BI343877
LOCUS       BI343877                596 bp    mRNA    linear   EST 30-JUL-2001
DEFINITION  372176 MARC 2PIG Sus scrofa cDNA 5', mRNA sequence.
ACCESSION   BI343877
VERSION     BI343877.1  GI:15037166
KEYWORDS    EST.
SOURCE      pig.
  ORGANISM  Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE   1  (bases 1 to 596)
  AUTHORS   Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Smith,T.P.L., Casas,E.,
            Stone,R.T., Heaton,M.P., Grosse,W.M., Bennett,G.A., Laegreid,W.W.
            and Keele,J.W.
  TITLE     Design and use of two pooled tissue normalized cDNA libraries for
            EST discovery in swine
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@email.marc.usda.gov
            Single pass sequencing. Bases called and alt_trimmed with phred
            v0.980904.e. Vector identified by cross_match with the -minscore 18
            and -minmatch 12 options.
            PCR PRimers
            FORWARD: AGGAAACAGCTATGACCAT
            BACKWARD: GTTTTCCCAGTCACGACG
            Plate: 115  row: O  column: 1
            Seq primer: ATTTAGGTGACACTATAG.
FEATURES             Location/Qualifiers
     source          1. .596
                     /organism="Sus scrofa"
                     /db_xref="taxon:9823"
                     /clone_lib="MARC 2PIG"
                     /tissue_type="pooled"
                     /lab_host="DH10B"
                     /note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
                     Library made from pooled tissue from testis, ovary,
                     endometrium, hypothalamus, pituitary, and placenta."
BASE COUNT      190 a    134 c    126 g    145 t      1 others
ORIGIN

  Query Match             21.2%;  Score 445;  DB 10;  Length 596;
  Best Local Similarity   89.0%;  Pred. No. 1.2e-113;
  Matches  492;  Conservative    0;  Mismatches   60;  Indels    1;  Gaps    1;

Qy   1550 ATCACCTTCTTCCGTCACCAGAGGGCGAGATCGCACCAGCCATCACATTGACCGTTAACA 1609
          |||||||||||||||||||||||||||||||||| ||||||||||||||  |||| || | ||
Db      1 ATCACCTTCTTCCGTCACCAGAGGGCGAGATTGCACCAGCCATCACGCTGACTGTCACCA 60

Qy   1610 TTTCGGGCAAGCTGGTGGATTTTGTCGTGACACACTTTGGGAACCACGAAGATGACCTCG 1669
          ||||   ||  |||||||| |||||||||| || ||||||||  ||||||||||||||||
Db     61 TTTCCAACAGACTGGTGGACTTTGTCGTGACCCATTTTGGGAATGACGAAGATGACCTCG 120

Qy   1670 ACAGGAAACTGCAGGCTATTGCTGTTTCAAAACTAC-TGAAAAGTAGCTCTAATCAAGTG 1728
          |||| ||||| || |||||||| || |||||||||| |||||| ||||||| | ||||||
Db    121 ACAGAAAACTTCAAGCTATTGCCGTCTCAAAACTACNTGAAAAATAGCTCTGAGCAAGTG 180

Qy   1729 ATATTTCTGGGATATATCACTTCAGCACCTGGCTCCAGAGATTATCTACAGCTCACTGAA 1788
```

```
          |||||||||||||| |||||||||  |||||| |||||||||||||| |||||||| ||||
Db    181 ATATTTCTGGGATACATCACTTCAAAACCTGGTTCCAGAGATTATCGACAGCTCATTGAA 240

Qy   1789 CATGGCAATGTGAAGGATATCGACAGCACTGATCATGACAGATGGTGTGAATACATTATG 1848
          |||||||||||||||||||||||||||||| || |  |||||||||||||||| ||||||
Db    241 CATGGCAATGTGAAGGATATCGACAGCACAGACCGAGACAGATGGTGTGAATATATTATG 300

Qy   1849 TATCGAGGGCTGATCAGGTTGGGTTATGCAAGAATCTCCCATGCTGAACTGAGTGATTCA 1908
          || ||||||||||||||  |||||||||||||||||||||| |||||||||||||||| ||
Db    301 TACCGAGGGCTGATCAGACTGGGTTATGCAAGAATCTCCCGTGCTGAACTGAGTGACTCT 360

Qy   1909 GAAATTCAGATGGCAAAATTTAGGATCCCTGATGACCCCACTAATTATAGAGACAACCAG 1968
          |||||||||||||| |||||||||||||||||||||||||  ||||||||| |||||||||
Db    361 GAAATTCAGATGGCCAAATTTAGGATCCCTGATGACCCCCTTAATTATAGGGACAACCAG 420

Qy   1969 AAAGTGGTCATAGACCACAGAGAAGTTTCTGAGAAAATTCATTTTAATCCCAGATTTGGA 2028
          || ||||||| |||||||||  ||||| ||||||| ||||||| |||||||| |||||
Db    421 AAGGTGGTCACAGACCACAGGAAAGTTCCTGAGAAGATTCATTTCAATCCCAGGTTTGGG 480

Qy   2029 TCCTACAAAGAAGGACACAATTATGAAAACAACCATCATTTTCATATGAATACTCCCAAA 2088
          ||||||||||||||||||||  ||| ||||| |||||||||| |||| ||||||||||
Db    481 TCCTACAAAGAAGGACACAATGCTGAGAACAATCATCATTTTCACATGAGTACTCCCAAA 540

Qy   2089 TACTTTTTATGAA 2101
          |||||||||||||
Db    541 TACTTTTTATGAA 553


RESULT 9
BG247983
LOCUS       BG247983                966 bp    mRNA    linear   EST 13-FEB-2001
DEFINITION  602359913F1 NCI_CGAP_Mam1 Mus musculus cDNA clone IMAGE:4488290 5',
            mRNA sequence.
ACCESSION   BG247983
VERSION     BG247983.1  GI:12757798
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 966)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Gilbert Smith, Ph.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10334  row: j  column: 03
            High quality sequence stop: 653.
FEATURES             Location/Qualifiers
     source          1. .966
                     /organism="Mus musculus"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:4488290"
                     /clone_lib="NCI_CGAP_Mam1"
                     /tissue_type="tumor, biopsy sample"
                     /dev_stage="10 months, virgin"
                     /lab_host="DH10B"
                     /note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
                     Site_2: NotI; Cloned unidirectionally.  Primer: Oligo dT.
                     Library constructed by Life Technologies. Investigator
                     providing samples: Gilbert Smith, NIH"
BASE COUNT      254 a    204 c    262 g    245 t      1 others
ORIGIN

  Query Match             21.0%;  Score 441.6;  DB 10;  Length 966;
  Best Local Similarity   84.1%;  Pred. No. 1.3e-112;
  Matches  556;  Conservative    0;  Mismatches  100;  Indels    5;  Gaps    5;

Qy    914 TTATTAACTCAGGGACAAACCCTGGGAAAACCATGACCATTGCCATGATATTTTATCTTC 973
          || | |||||||||| |||| | ||  || ||||| ||| || ||||||  ||||| ||
Db      1 TTGTGAACTCAGGGAAAAACTCCGGAGAAGCCATGGCCACTGGCATGATCCTTTATGTTT 60

Qy    974 TAGAAATATTTTTCTGTGCCTGGTGCACAGCTTTTAAGTTTGTCCCAGGAGGTGTCTACG 1033
          || ||| |||||||||||||||||||||||| |||||| |||| || ||||||||||||
Db     61 TACAAACATTTTTCTGTGCCTGGTGCACAGCGTTTAAG-TTGTTCCCGGAGGTGTCTACG 119

Qy   1034 CTAGAGAAAGATCAGATGTGCTTTTGGGGACAATGATGTTAATTATCGGGCTGAATATGC 1093
          |||||||||||||| |||||||||||||||||| || ||| ||||||| ||  ||| |||||
Db    120 CTAGAGAAAGATCTGATGTGCTTTTGGGGACCATAATGGTAATTATTGGATTGAGTATGC 179

Qy   1094 TATTTGGTCCTAAGAAAAACCTTGACTTGCTTCTTCAAACAAAAAACAGTTCTAAAGTGC 1153
          | ||||| |||||||||||||||||||| |||||||||||||||||||| ||||| ||
Db    180 TGTTTGGACCTAAGAAAAACCTTGACTTTCTTCTTCAAACAAAAAACAGTCCTAAAACGC 239

Qy   1154 TTTTCAGAAAGAGTGAAAAATACATGAAACTTTTTCTGTGGCTGCTTGTTGGTGTGGGAT 1213
          |||| |||  |||||||||||||||||||  |||||||||| |||||||||||| |
Db    240 TTTTGAGATGCAGTGAAAAATACATGAAACTCATTCTGTGGCTG-TTGTTGGTGTGGGGT 298

Qy   1214 TGTTGGGATTAGGACTACGGCATAAAGCCTATGAGAGAAAACTGGGCAAAGTGGCACCAA 1273
          |  |||||||||| |||||||| | | |||| |||||  |  | |||  || |||||||
Db    299 TACTGGGATTAGGGCTACGGCACAGAACCTACGAGAGGCAGTTAGGCCGAGGGGCACCAG 358

Qy   1274 CCAAAGAGGTCTCTGCTGCCATCTGGCCTTTCAGGTTTGGATATGACAATGAAGGGTGGT 1333
          | | || |||||| |||||||||||||||| ||| ||||||||||||||||||| |||
Db    359 CGACAGTGGTCTCGGCTGCCATCTGGCCTTTTAGG-TTGGATATGACAATGAAGGCTGGC 417

Qy   1334 CTAGTCTAGAAAGATCAGCTCACCTGCTCAATGAAACAGGTGCAGATTTCATAACAATTT 1393
          | | |||||| || || ||||| |||||||| || |||||||||||||| || || ||||
Db    418 CCAATCTAGAGAGGTCTGCTCAACTGCTCAAGGAGACAGGTGCAGATTTTATCACCATTT 477

Qy   1394 TGGAGAGTGATGCTTCTAAGCCCTATATGGGGAACAATGACTTAACCATGTGGCTAGGGG 1453
          ||||||| |||||||| |||||||||| |||||||| |||||||| |||||||| ||||
Db    478 TGGAGAGCGATGCTTCCAAGCCCTATATAGGGAACAACGACTTAACAATGTGGCTGGGGG 537

Qy   1454 AAAAGTTGGGTTTCTATACAGACTTTGGTCCAAGCACAAGGTATCACACTTGGGGGATTA 1513
          | ||| |||||||||| |||||||||| |||||||| ||| |||||||| ||||||||||
Db    538 AGAAGCTGGGTTTCTACACAGACTTTGGGCCAAGCACCAGG-ATCACACCTGGGGGATTA 596

Qy   1514 TGGCTTTGTCAAGATACCCAATTGTGAAATCTGAGCATCACCTTCTTCCGTCACCAGAGG 1573
          |||   |||| | || || ||||| | ||| || |||||||||||||||| ||||||
Db    597 TGGTGCTGTCTCGGTA-CCGATTGTCAGATCGGAACATCACCTTCTTCCGTCGCCAGAGN 655

Qy   1574 G 1574
          |
Db    656 G 656


RESULT 10
AA710301
LOCUS       AA710301                597 bp    mRNA    linear   EST 24-DEC-1997
DEFINITION  vt53b05.r1 Barstead mouse irradiated colon MPLRB7 Mus musculus cDNA
            clone IMAGE:1166769 5' similar to TR:P78781 P78781 FISSION YEAST ;,
            mRNA sequence.
ACCESSION   AA710301
VERSION     AA710301.1  GI:2720219
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 597)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
```

```
COMMENT     Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:632681
            Seq primer: -28m13 rev2 ET from Amersham
            High quality sequence stop: 509.
FEATURES             Location/Qualifiers
     source          1. .597
                     /organism="Mus musculus"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:1166769"
                     /clone_lib="Barstead mouse irradiated colon MPLRB7"
                     /dev_stage="8 weeks"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: EcoRI; Site_2: NotI; Tissue obtained
                     from 8 week old mouse.  Colon was harvested 72 hours after
                     irradiation with 1400 Gys.  1st strand cDNA was primed
                     with  a Not I - oligo(dT) primer
                     [5'TGTTACGAATCTGAAGTGGGAGCGGCCGCCCTTTTTTTTTTTTTTTTTTTTTTTT
                     T 3'];  double-stranded cDNA was ligated to Eco RI
                     adaptors [AATTCGGATCCTTG],  digested with Not I and cloned
                     into the Not I and Eco RI sites of  the modified pT7T3
                     vector. Library constructed by Bob Barstead.  "
BASE COUNT      174 a    141 c    144 g    138 t
ORIGIN

  Query Match             20.8%;  Score 436.6;  DB 9;  Length 597;
  Best Local Similarity   84.6%;  Pred. No. 2.7e-111;
  Matches  490;  Conservative    0;  Mismatches   89;  Indels    0;  Gaps    0;

Qy   1469 ATACAGACTTTGGTCCAAGCACAAGGTATCACACTTGGGGGATTATGGCTTTGTCAAGAT 1528
           || |||||||||| |||||||| ||| ||||||| ||||||||||||    |||| | |
Db     19 ATGCAGACTTTGGGCCAAGCACCAGGGATCACACCTGGGGGATTATGGTGCTGTCTCGGT 78

Qy   1529 ACCCAATTGTGAAATCTGAGCATCACCTTCTTCCGTCACCAGAGGGCGAGATCGCACCAG 1588
           |||| ||||| | ||| || ||||||||||||||||| |||||||||    |||||||||
Db     79 ACCCGATTGTCAGATCGGAACATCACCTTCTTCCGTCGCCAGAGGGCAGAGTCGCACCAG 138

Qy   1589 CCATCACATTGACCGTTAACATTTCGGGCAAGCTGGTGGATTTTGTCGTGACACACTTTG 1648
           |||| ||  |||| |||||| | ||   ||  ||||||||||||| ||||||||||||
Db    139 CCATAACCATGACAGTTAACGTCTCCAACAGACTGGTGGATTTTGTGGTGACACACTTTG 198

Qy   1649 GGAACCACGAAGATGACCTCGACAGGAAACTGCAGGCTATTGCTGTTTCAAAACTACTGA 1708
           |||| || |||||||||| |||||||| || |||||||||| |||||||||| ||||
Db    199 GGAATCATGAAGATGACCTTGACAGGAAGCTACAGGCTATTGCAGTTTCAAAACTGCTGA 258

Qy   1709 AAAGTAGCTCTAATCAAGTGATATTTCTGGGATATATCACTTCAGCACCTGGCTCCAGAG 1768
           ||| | | || |||||||||||||||||||| ||||||||||||  ||||| |||||||
Db    259 AAAATTGTTCAAATCAAGTGATATTTCTGGGTTATATCACTTCAGAGCCTGGTTCCAGAG 318

Qy   1769 ATTATCTACAGCTCACTGAACATGGCAATGTGAAGGATATCGACAGCACTGATCATGACA 1828
           ||||  |||| |||||  |||||||||| |||||||||| |||||| | |||   ||||
Db    319 ATTACATACAACTCACAAAACATGGCAACGTGAAGGATATTGACAGCTCAGATGGAGACA 378

Qy   1829 GATGGTGTGAATACATTATGTATCGAGGGCTGATCAGGTTGGGTTATGCAAGAATCTCCC 1888
           |||||||||| ||||| ||||| || || |||||||||||||||| |||||||||||| |
Db    379 GATGGTGTGAGTACATCATGTACCGGGGTCTGATCAGGTTGGGCTATGCAAGAATCTCTC 438

Qy   1889 ATGCTGAACTGAGTGATTCAGAAATTCAGATGGCAAAATTTAGGATCCCTGATGACCCCA 1948
           |||| |||||||||| || || |||||||||| ||||| |||||||| ||||| |||
Db    439 ATGCAGAACTGAGTGACTCTGAGATTCAGATGGCCAAATTCAGGATCCCGGATGATCCCG 498

Qy   1949 CTAATTATAGAGACAACCAGAAAGTGGTCATAGACCACAGAGAAGTTTCTGAGAAAATTC 2008
           | ||||| |||||||||||||||| |||||||||| ||| |||| |  |||| ||||
Db    499 CCAATTACAGAGACAACCAGAAAGTCGTCATAGACCACCGAGGAGTTCCCAAGAACATTC 558

Qy   2009 ATTTTAATCCCAGATTTGGATCCTACAAAGAAGGACACA 2047
           | || || |||||||||| ||||||||||||||||||
Db    559 ACTTCAACCCCAGATTTGGCTCCTACAAAGAAGGACACA 597


RESULT 11
AW235545/c
LOCUS       AW235545               631 bp    mRNA    linear   EST 13-DEC-1999
DEFINITION  xn19e05.x1 NCI_CGAP_Kid11 Homo sapiens cDNA clone IMAGE:2694176 3'
            similar to SW:YCQ7_YEAST P25618 HYPOTHETICAL 107.9 KD PROTEIN IN
            POL4-SRD1 INTERGENIC REGION. ;, mRNA sequence.
ACCESSION   AW235545
VERSION     AW235545.1  GI:6567934
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 631)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
            cDNA Library Preparation: M. Bento Soares, Ph.D.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Seq primer: -40UP from Gibco
            High quality sequence stop: 441.
FEATURES             Location/Qualifiers
     source          1. .631
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2694176"
                     /clone_lib="NCI_CGAP_Kid11"
                     /lab_host="DH10B"
                     /note="Organ: kidney; Vector: pT7T3D-Pac (Pharmacia) with
                     a modified polylinker; Site_1: Not I; Site_2: Eco RI;
                     Plasmid DNA from the normalized library NCI_CGAP_Kid3 was
                     prepared, and ss circles were made in vitro. Following HAP
                     purification, this DNA was used as tracer in a subtractive
                     hybridization reaction. The driver was PCR-amplified cDNAs
                     from a pool of 5,000 clones made from the same library
                     (cloneIDs 1322376-1323911, 1456007-1456775, and
                     1500552-1502855). Subtraction by Bento Soares and M.
                     Fatima Bonaldo.  "
BASE COUNT      168 a    116 c    101 g    245 t      1 others
ORIGIN

  Query Match             19.9%;  Score 418;  DB 9;  Length 631;
  Best Local Similarity   97.5%;  Pred. No. 4.7e-106;
  Matches  424;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy   1669 GACAGGAAACTGCAGGCTATTGCTGTTTCAAAACTACTGAAAAGTAGCTCTAATCAAGTG 1728
           |||||||| |||||||||||||||||||| |  ||||||| |||| ||||||||||
Db    631 GACAGGAAAATGCAGGCTATTGCTGTTTCAAAATTCTTGAAAAGAAGCTTTAATCAAGTG 572

Qy   1729 ATATTTCTGGGATATATCACTTCAGCACCTGGCTCCAGAGATTATCTACAGCTCACTGAA 1788
           |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    571 ATATTTNTGGGATATATCACTTCAGCACCTGGCTCCAGAGATTATCTACAGCTCACTGAA 512

Qy   1789 CATGGCAATGTGAAGGATATCGACAGCACTGATCATGACAGATGGTGTGAATACATTATG 1848
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    511 CATGGCAATGTGAAGGATATCGACAGCACTGATCATGACAGATGGTGTGAATACATTATG 452

Qy   1849 TATCGAGGGCTGATCAGGTTGGGTTATGCAAGAATCTCCCATGCTGAACTGAGTGATTCA 1908
```

```
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    451 TATCGAGGGCTGATCAGGTTGGGTTATGCAAGAATCTCCCATGCTGAACTGAGTGATTCA 392

Qy   1909 GAAATTCAGATGGCAAAATTTAGGATCCCTGATGACCCCACTAATTATAGAGACAACCAG 1968
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    391 GAAATTCAGATGGCAAAATTTAGGATCCCTGATGACCCCACTAATTATAGAGACAACCAG 332

Qy   1969 AAAGTGGTCATAGACCACAGAGAAGTTTCTGAGAAAATTCATTTTAATCCCAGATTTGGA 2028
          |||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||
Db    331 AAAGTGGTCATAGACCACAGAGAAGTTTTTGAGAAAATTCATTTTAATCCCAGATTTGGA 272

Qy   2029 TCCTACAAAGAAGGACACAATTATGAAAACAACCATCATTTTCATATGAATACTCCCAAA 2088
          |||||||||||||||| |||||||||||||| |||| ||||||||||||||||||||||||
Db    271 TCCTACAAAGAAGGACCCAATTATGAAAACACCCATAATTTTCATATGAATACTCCCAAA 212

Qy   2089 TACTTTTTATGAAAC 2103
          |||||||||||||||
Db    211 TACTTTTTATGAAAC 197
```

RESULT 12
BE180939

LOCUS	BE180939 528 bp mRNA linear EST 22-JUN-2000
DEFINITION	PM3-HT0627-270300-001-e03 HT0627 Homo sapiens cDNA, mRNA sequence.
ACCESSION	BE180939
VERSION	BE180939.1 GI:8660115
KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE	1 (bases 1 to 528)
AUTHORS	Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.
TITLE	Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE	20202663
COMMENT	Contact: Simpson A.J.G. Laboratory of Cancer Genetics Ludwig Institute for Cancer Research Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil Tel: +55-11-2704922 Fax: +55-11-2707001 Email: asimpson@ludwig.org.br This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=&t2=PM3-HT0627-270300-001-e03&t3=2000-03-27&t4=1) Seq primer: puc 18 forward High quality sequence start: 2 High quality sequence stop: 475.

FEATURES	Location/Qualifiers
source	1. .528 /organism="Homo sapiens" /db_xref="taxon:9606" /clone_lib="HT0627" /dev_stage="Adult" /note="Organ: head_neck; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 151 a 114 c 131 g 132 t
ORIGIN

```
 Query Match           19.6%;  Score 411.8;  DB 9;  Length 528;
 Best Local Similarity  95.1%;  Pred. No. 2.4e-104;
 Matches  490;  Conservative    0;  Mismatches   17;  Indels    8;  Gaps    6;

Qy   1218 GGGATTAGGACTACGGCATAAAGCCTATGAGAGAAAACTGGGCAAAGTGGCACCAACCAA 1277
          ||||||||||||||    ||    |||||||||||||||||| ||||||||||||||||
Db      3 GGGATTAGGACTACG--CTAGTTACTATGAGAGAAAACTGGGC-AAGTGGCACCAACCAA 59

Qy   1278 AGAGGTCTCTGCTGCCATCTGGCCTTTCAGGTTTGGATATGACAATGAAGGGTGGTCTAG 1337
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     60 AGAGGTCTCTGCTGCCATCTGGCCTTTCAGGTTTGGATATGACAATGAAGGGTGGTCTAG 119

Qy   1338 TCTAGAAAGATCAGCTCACCTGCTCAATGAAACAGGTGCAGATTTCATAACAATTTTGGA 1397
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    120 TCTAGAAAGATCAGCTCACCTGCTCAATGAAACAGGTGCAGATTTCATAACAATTTTGGA 179

Qy   1398 GAGTGATGCTTCTAAGCCCTATATGGGGAACAATGACTTAACCATGTGGCTAGGGGAAAA 1457
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    180 GAGTGATGCTTCTAAGCCCTATATGGGGAACAATGACTTAACCATGTGGCTAGGGGAAAA 239

Qy   1458 GTTGGGTTTCTATACAGACTTTGGTCCAAGCACAAGGTATCACACTTGGGGGATTATGGC 1517
          |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db    240 GTTGGGTTTCTATACAGACTTTGGTCCAAGCACAAGGTATCACACTTGGCGGATTATGGC 299

Qy   1518 TTTGTCAAGATACCCAATTGTGAAATCTGAGCATCACCTTCTTCCGTCACCAGAGGGCGA 1577
          ||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    300 TTTGTCAAGGTACCCAATTGTGAAATCTGAGCATCACCTTCTTCCGTCACCAGAGGGCGA 359

Qy   1578 GATCGCACCAGCCATCACATTGA-CCGTTAACATTTCGGGCAA--GCTGGTGGATTTTGT 1634
          |||||||||||||||||||||| ||||||||||||||||||     | ||||| |||
Db    360 GATCGCACCAGCCATCACATTGACCCGTTAACATTTCGGGCAACGCTGGTGGGATTATGT 419

Qy   1635 CGTGACACACTTTGGGAACCA-CGAAGATGACCTCGACAGGAAACTGC-AGGCTATTGCT 1692
          |||||||||||||||||||| |||||||||||||||||||||||||| ||||||||| |
Db    420 CGTGACACACTTTGGGAACCACCGAAGATGACCTCGACAGGAAACTGCAAGGCTATTGGT 479

Qy   1693 GTTTCAAAACTACTGAAAAGTAGCTCTAATCAAGT 1727
          ||| || ||||||||||||||||||||||||| ||
Db    480 GTTACACAACTACTGAAAAGTAGCTCTAATCAGGT 514
```

RESULT 13
BI853011

LOCUS	BI853011 974 bp mRNA linear EST 10-OCT-2001
DEFINITION	603379834F1 NIH_CGAP_Mam3 Mus musculus cDNA clone IMAGE:5389784 5', mRNA sequence.
ACCESSION	BI853011
VERSION	BI853011.1 GI:15993758
KEYWORDS	EST.
SOURCE	house mouse.
ORGANISM	Mus musculus Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE	1 (bases 1 to 974)
AUTHORS	NIH-MGC http://mgc.nci.nih.gov/.
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL	Unpublished (1999)
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgapbs-r@mail.nih.gov Tissue Procurement: Lothar Hennighausen Ph.D., Chu-Xia Deng Ph.D. cDNA Library Preparation: Life Technologies, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: LLAM11993 row: d column: 09 High quality sequence stop: 765.

FEATURES	Location/Qualifiers
source	1. .974 /organism="Mus musculus" /db_xref="taxon:10090" /clone="IMAGE:5389784"

```
                /clone_lib="NIH_CGAP_Mam3"
                /tissue_type="tumor, gross tissue"
                /lab_host="DH10B"
                /note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: NotI;
                Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                Average insert 2 kb. Library constructed by Life
                Technologies, catalog #12017-018. Investigators providing
                samples: Lothar Hennighausen/Chu-Xia Deng, NIH Reference
                for transgenic model: Xu et al., Nature Genetics 22, 37-43
                (1999). Note: this is a NCI_CGAP Library."
BASE COUNT      198 a    272 c    274 g    230 t
ORIGIN

  Query Match             19.5%;  Score 409.4;  DB 10;  Length 974;
  Best Local Similarity   78.1%;  Pred. No. 1.5e-103;
  Matches  505;  Conservative    0;  Mismatches  141;  Indels    1;  Gaps    1;

Qy       1 ATGACCTCGCTGTGGAGAGAAATCCTCTTGGAGTCGCTGCTGGGATGTGTTTCTTGGTCT 60
             |||  |    |||||||||| |||   | |||||  |  |||||||  |  ||  || ||||||
Db     123 ATGCCAGGCCTGTGGAGAGCAATTGCCCTGGAGACCTTGCTGGGCTACGTCTCCTGGTCT 182

Qy      61 CTCTACCATGACCTGGGACCGATGATCTATTACTTTCCTTTGCAAACACTAGAACTCACT 120
           ||||||||  |  |||  | ||  |||||||||||  |||||  |||||||  ||  ||  |||||
Db     183 CTCTACCACGGACTGAGCCCCATGATCTATTATTTTCCCCTGCAAACGCTGGAGCTCACA 242

Qy     121 GGGCTTGAAGGTTTTAGTATAGCATTTCTTTCTCCAATATTCCTAACAATTACTCCTTTC 180
           |||||  |||   |||  ||  ||||  |||||  |||||  |||  ||  ||||||||  |  ||  ||
Db     243 GGGCTGGAATTCTTTTGTGTAGCCTTTCTCTCTCCGATACTCTTAACAATTCCCCCACTC 302

Qy     181 TGGAAATTGGTTAACAAGAAGTGGATGCTAACCCTGCTGAGGATAATCACTATTGGCAGC 240
           ||||||  |  ||  ||  ||||||||||  ||  |  ||||||||||||  |||||  |  ||||||
Db     303 TGGAAACTCGTCAATAAGAAGTGGACACTCTCTCTGCTGAGGATCGTCACTGTAGGCAGC 362

Qy     241 ATAGCCTCCTTCCAGGCTCCAAATGCCAAACTTCGACTGATGGTTCTTGCGCTTGGGGTG 300
           |||||  |||||  ||||||||||||||  ||  ||||||||||||||  ||  ||  |||||  |||
Db     363 ATAGCTTCCTTTGAGGCTCCAAATGCTAAGCTTCGACTGATGGTCCTCGCCCTTGGTGTG 422

Qy     301 TCTTCCTCACTGATAGTGCAAGCTGTGACTTGGTGGTCGGGAAGTCATTTGCAAAGGTAC 360
           |||||  ||  ||||||||||||  |||||||  ||||||||  ||||||    ||||||||||||
Db     423 TCTTCTTCCCTGATAGTGCAAACTGTGACCTGGTGGTCAGGAAGTGGATTGCAAAGGTAC 482

Qy     361 CTCAGAATTTGGGGATTCATTTTAGGACAGATTGTTCTTGTTGTTCTACGCATATGGTAT 420
           ||||  |||  ||||||||||||  |  |||||  |  |||||  ||||  ||  ||  ||||||||
Db     483 CTCAAAATCTGGGGATTCATTCTGGGACACGTCCTTCTTCTTGTCCTCCGAATATGGTAC 542

Qy     421 ACTTCACTAAACCCAATCTGGAGTTATCAGATGTCCAACAAAGTGATACTGACATTAAGT 480
           ||  |||  ||||||||||  |||||  ||||||||||||||||  |||  ||  ||||||||  |||
Db     543 ACATCATTAAACCCAATTTGGAGCTATCAGATGTCCAACAGAGTAATTCTGACATTGAGT 602

Qy     481 GCCATAGCCACACTTGATCGTATTGGCACAGATGGTGACTGCAGTAAACCTGAAGAAAAG 540
           |||  ||||    |||||  ||  ||||||||||||||||||  |||  ||  |||||  |  |||
Db     603 GCCGTAGCTGTGCTTGACCGAATTGGCACAGATGGTGACTACAGAAACCCTGAGGGGAAG 662

Qy     541 AAGACTGGTGAGGTAGCCACGGGG-ATGGCCTCTAGACCCAACTGGCTGCTGGCAGGGGC 599
           |||  |  |||||||  |||||||||  |  |  |||||  |  |  |  ||||||  ||  ||||  ||
Db     663 AAGCCCCGTGAGGTTGCCACGGGGCAGGACCTCTCTATCTAGTTGGCTGTTGCCAGGCGC 722

Qy     600 TGCTTTTGGTAGCCTTGTGTTCCTCACCCACTGGGTTTTTGGAGAAG 646
           |||||  ||    |      |||||  ||  ||||||  |  ||  |||||||
Db     723 CGCTTTGGGAGCCTCCTGTTTCCTAACTCACTGGATCTTCGGAGAAG 769


RESULT 14
AI962905/c
LOCUS       AI962905                617 bp    mRNA    linear   EST 20-AUG-1999
DEFINITION  wt24f04.x1 NCI_CGAP_Ut1 Homo sapiens cDNA clone IMAGE:2508415 3'
            similar to TR:P78781 P78781 FISSION YEAST ;, mRNA sequence.
ACCESSION   AI962905
VERSION     AI962905.1  GI:5755618
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 617)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Seq primer: -40UP from Gibco
            High quality sequence stop: 407.
FEATURES             Location/Qualifiers
     source          1. .617
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2508415"
                     /clone_lib="NCI_CGAP_Ut1"
                     /tissue_type="well-differentiated endometrial
                     adenocarcinoma, 7 pooled tumors"
                     /lab_host="DH10B"
                     /note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: SalI;
                     Site_2: NotI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 1.75 kb. Life Technologies catalog #:
                     11538-014"
BASE COUNT      159 a    109 c    102 g    246 t      1 others
ORIGIN

  Query Match             18.8%;  Score 395.8;  DB 9;  Length 617;
  Best Local Similarity   99.0%;  Pred. No. 7.9e-100;
  Matches  408;  Conservative    0;  Mismatches    3;  Indels    1;  Gaps    1;

Qy    1693 GTTTCAAAACTACTGAAAAGTAGCTCTAATCAAGTGATATTTCTGGGATATATCACTTCA 1752
           ||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     617 GTTTCAAAACTACTGNAAAGTAGCTCTAATCAAGTGATATTTCTGGGATATATCACTTCA 558

Qy    1753 GCACCTGGCTCCAGAGATTATCTACAGCTCACTGAACATGGCAATGTGAAGGATATCGAC 1812
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     557 GCACCTGGCTCCAGAGATTATCTACAGCTCACTGAACATGGCAATGTGAAGGATATCGAC 498

Qy    1813 AGCACTGATCATGACAGATGGTGTGAATACATTATGTATCGAGGGCTGATCAGGTTGGGT 1872
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     497 AGCACTGATCATGACAGATGGTGTGAATACATTATGTATCGAGGGCTGATCAGGTTGGGT 438

Qy    1873 TATGCAAGAATCTCCCATGCTGAACTGAGTGATTCAGAAATTCAGATGGCAAAATTTAGG 1932
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     437 TATGCAAGAATCTCCCATGCTGAACTGAGTGATTCAGAAATTCAGATGGCAAAATTTAGG 378

Qy    1933 ATCCCTGATGACCCCACTAATTATAGAGACAACCAGAAAGTGGTCATAGACCACAGAGAA 1992
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     377 ATCCCTGATGACCCCACTAATTATAGAGACAACCAGAAAGTGGTCATAGACCACAGAGAA 318

Qy    1993 GTTTCTGAGAAAATTCATTTTAATCCCAGATTTGGATCCTACAAAGAAGGACACAATTAT 2052
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db     317 GTTTCTGAGAAAATTCATTTTAATCCCAGATTTGGATCCTACAAAGAAGGACCCAATTAT 258

Qy    2053 GAAAACAACCATCATTTTCATATGAATACT-CCCAAATACTTTTTATGAAAC 2103
           |||||||||||| |||||||||||||||| ||||||||||||||||||||
Db     257 GAAAACAACCATAATTTTCATATGAATACTCCCCAAATACTTTTTATGAAAC 206


RESULT 15
BB628092
LOCUS       BB628092                656 bp    mRNA    linear   EST 31-AUG-2001
DEFINITION  BB628092 RIKEN full-length enriched, adult male urinary bladder Mus
            musculus cDNA clone 9530064O17 5', mRNA sequence.
ACCESSION   BB628092
```

VERSION BB628092.1 GI:15399293
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 656)
AUTHORS Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T., Muramatsu,M. and Hayashizaki,Y.
TITLE RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
JOURNAL Unpublished (2001)
COMMENT Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. . 10 (10), 1617-1630 (2000)
wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. . 10 (11), 1757-1771 (2000)
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara,Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
Yamanaka,I., Kiyosawa,H., Kondo,S., Saito,T., Shinagawa,A., Aizawa,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K., Arakawa,T., Ishii,Y. and Hayashizaki,Y.
Mapping of 19032 mouse cDNAs on mouse chromosomes. J. Struct. Func. Genomics 2 pre, L72-L86 (2001
Please visit our web site (http://genome.gsc.riken.go.jp) for further details.
e mouse tissues.
FEATURES Location/Qualifiers
source 1. .656
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="9530064O17"
/clone_lib="RIKEN full-length enriched, adult male urinary bladder"
/sex="male"
/tissue_type="urinary bladder"
/dev_stage="adult"
/lab_host="DH10B"
/note="Site_1: SalI; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 20.0 and subtraction to Rot = 370.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC 3']. cDNA was cleaved with XhoI and BamHI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I."

BASE COUNT 135 a 183 c 173 g 164 t 1 others
ORIGIN

Query Match 16.8%; Score 354.2; DB 9; Length 656;
Best Local Similarity 80.0%; Pred. No. 3.9e-88;
Matches 416; Conservative 0; Mismatches 104; Indels 0; Gaps 0;

```
Qy      1 ATGACCTCGCTGTGGAGAGAAATCCTCTTGGAGTCGCTGCTGGGATGTGTTTCTTGGTCT 60
          ||| |    |||||||||| |||   | ||||| |  |||||||  |  || || ||||||
Db    136 ATGCCAGGCCTGTGGAGAGCAATTGCCCTGGAGACCTTGCTGGGCTACGTCTCCTGGTCT 195

Qy     61 CTCTACCATGACCTGGGACCGATGATCTATTACTTTCCTTTGCAAACACTAGAACTCACT 120
          |||||||| |  ||| | || |||||||||| |||||  |||||| || || |||||
Db    196 CTCTACCACGGACTGAGCCCCATGATCTATTATTTTCCCCTGCAAACGCTGGAGCTCACA 255

Qy    121 GGGCTTGAAGGTTTTAGTATAGCATTTCTTTCTCCAATATTCCTAACAATTACTCCTTTC 180
          ||||| |||   ||| || |||| ||||| ||||| ||| || |||||||| | ||  ||
Db    256 GGGCTGGAATTCTTTTGTGTAGCCTTTCTCTCTCCGATACTCTTAACAATTCCCCCACTC 315

Qy    181 TGGAAATTGGTTAACAAGAAGTGGATGCTAACCCTGCTGAGGATAATCACTATTGGCAGC 240
          |||||| | || || |||||||||  ||  | ||||||||||  ||||| | ||||||
Db    316 TGGAAACTCGTCAATAAGAAGTGGACACTCTCTCTGCTGAGGATCGTCACTGTAGGCAGC 375

Qy    241 ATAGCCTCCTTCCAGGCTCCAAATGCCAAACTTCGACTGATGGTTCTTGCGCTTGGGGTG 300
          ||||| |||||  |||||||||||||| || |||||||||||||| || || ||||| |||
Db    376 ATAGCTTCCTTTGAGGCTCCAAATGCTAAGCTTCGACTGATGGTCCTCGCCCTTGGTGTG 435

Qy    301 TCTTCCTCACTGATAGTGCAAGCTGTGACTTGGTGGTCGGGAAGTCATTTGCAAAGGTAC 360
          ||||| || |||||||||||| ||||||| |||||||| ||||||   ||||||||||||
Db    436 TCTTCTTCCCTGATAGTGCAAACTGTGACCTGGTGGTCAGGAAGTGGATTGCAAAGGTAC 495

Qy    361 CTCAGAATTTGGGGATTCATTTTAGGACAGATTGTTCTTGTTGTTCTACGCATATGGTAT 420
          |||| ||| |||||||||||| | |||||  |  ||||| |||| || || |||||||||
Db    496 CTCANAATCTGGGGATTCATTCTGGGACACGTCCTTCTTCTTGTCCTCCGAATATGGTAC 555

Qy    421 ACTTCACTAAACCCAATCTGGAGTTATCAGATGTCCAACAAAGTGATACTGACATTAAGT 480
          || ||| |||||||||| ||||| |||||||||||||||| ||| || ||||||| |||
Db    556 ACATCATTAAACCCAATTTGGAGCTATCAGATGTCCAACAGAGTAATTCTGACATTGAGT 615

Qy    481 GCCATAGCCACACTTGATCGTATTGGCACAGATGGTGACT 520
          ||| ||||    ||||| || ||||||||||||||||||
Db    616 GCCGTAGCTGTGCTTGACCGAATTGGCACAGATGGTGACT 655
```

Search completed: October 21, 2002, 21:55:24
Job time : 2493 secs

```
RESULT 1
Q9H720
ID   Q9H720      PRELIMINARY;      PRT;   699 AA.
AC   Q9H720;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   CDNA: FLJ21511 fis, clone COL05748.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=COLON;
RA   Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
RA   Okitani R., Ota T., Suzuki Y., Obayashi M., Nishi T., Shibahara T.,
RA   Tanaka T., Nakamura Y., Isogai T., Sugano S.;
RT   "NEDO human cDNA sequencing project.";
RL   Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AK025164; BAB15080.1; -.
SQ   SEQUENCE   699 AA;  78565 MW;  2F839563189A0523 CRC64;

  Query Match           99.8%;  Score 3713;  DB 4;  Length 699;
  Best Local Similarity 99.9%;  Pred. No. 3.6e-263;
  Matches 698;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 MTSLWREILLESLLGCVSWSLYHDLGPMIYYFPLQTLELTGLEGFSIAFLSPIFLTITPF 60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MTSLWREILLESLLGCVSWSLYHDLGPMIYYFPLQTLELTGLEGFSIAFLSPIFLTITPF 60

Qy      61 WKLVNKKWMLTLLRIITIGSIASFQAPNAKLRLMVLALGVSSSLIVQAVTWWSGSHLQRY 120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 WKLVNKKWMLTLLRIITIGSIASFQAPNAKLRLMVLALGVSSSLIVQAVTWWSGSHLQRY 120

Qy     121 LRIWGFILGQIVLVVLRIWYTSLNPIWSYQMSNKVILTLSAIATLDRIGTDGDCSKPEEK 180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 LRIWGFILGQIVLVVLRIWYTSLNPIWSYQMSNKVILTLSAIATLDRIGTDGDCSKPEEK 180

Qy     181 KTGEVATGMASRPNWLLAGAAFGSLVFLTHWVFGEVSLVSRWAVSGHPHPGPDPNPFGGA 240
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 KTGEVATGMASRPNWLLAGAAFGSLVFLTHWVFGEVSLVSRWAVSGHPHPGPDPNPFGGA 240

Qy     241 VLLCLASGLMLPSCLWFRGTGLIWWVTGTASAAGLLYLHTWAAAVSGCVFAIFTASMWPQ 300
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 VLLCLASGLMLPSCLWFRGTGLIWWVTGTASAAGLLYLHTWAAAVSGCVFAIFTASMWPQ 300

Qy     301 TLGHLINSGTNPGKTMTIAMIFYLLEIFFCAWCTAFKFVPGGVYARERSDVLLGTMMLII 360
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 TLGHLINSGTNPGKTMTIAMIFYLLEIFFCAWCTAFKFVPGGVYARERSDVLLGTMMLII 360

Qy     361 GLNMLFGPKKNLDLLLQTKNSSKVLFRKSEKYMKLFLWLLVGVGLLGLGLRHKAYERKLG 420
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 GLNMLFGPKKNLDLLLQTKNSSKVLFRKSEKYMKLFLWLLVGVGLLGLGLRHKAYERKLG 420

Qy     421 KVAPTKEVSAAIWPFRFGYDNEGWSSLERSAHLLNETGADFITILESDASKPYMGNNDLT 480
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 KVAPTKEVSAAIWPFRFGYDNEGWSSLERSAHLLNETGADFITILESDASKPYMGNNDLT 480

Qy     481 MWLGEKLGFYTDFGPSTRYHTWGIMALSRYPIVKSEHHLLPSPEGEIAPAITLTVNISGK 540
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 MWLGEKLGFYTDFGPSTRYHTWGIMALSRYPIVKSEHHLLPSPEGEIAPAITLTVNISGK 540

Qy     541 LVDFVVTHFGNHEDDLDRKLQAIAVSKLLKSSSNQVIFLGYITSAPGSRDYLQLTEHGNV 600
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 LVDFVVTHFGNHEDDLDRKLQAIAVSKLLKSSSNQVIFLGYITSAPGSRDYLQLTEHGNV 600

Qy     601 KDIDSTDHDRWCEYIMYRGLIRLGYARISHAELSDSEIQMAKFRIPDDPTNYRDNQKVVI 660
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 KDIDSTDHDRWCEYIMYRGLIRLGYARISHAELSDSEIQMAKFRIPDDPTNYRDNQKVVI 660

Qy     661 DHREVSEKIHFNPRFGSYKEGHNYENNHHFHMNTPKYFL 699
          |||||||||||||||||||||||||||||||||||||||||
Db     661 DHREVSEKIHFNPRFGSYKEGHNYENNHNFHMNTPKYFL 699
```

GenCore version 5.1.3

OM protein - protein search, using sw model

Run on: October 21, 2002, 21:56:53 ; Search time 63 Seconds
(without alignments)
1919.420 Million cell updates/sec

Title: US-09-809-638-2
Perfect score: 3720
Sequence: 1 MTSLWREILLESLLGCVSWS.....EGHNYENNHHFHMNTPKYFL 699

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_19:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	3713	99.8	699	4	Q9H720	Q9h720 homo sapien
2	3125	84.0	699	11	Q91YL7	Q91yl7 mus musculu
3	639	17.2	425	3	Q9P6M4	Q9p6m4 schizosacch
4	625.5	16.8	401	3	P78781	P78781 schizosacch
5	256	6.9	688	3	Q9HDZ2	Q9hdz2 schizosacch
6	141.5	3.8	279	2	Q9S1Z5	Q9s1z5 streptomyce
7	132	3.5	583	8	Q950T2	Q950t2 hyaloraphid
8	128.5	3.5	396	2	P95556	P95556 pseudomonas
9	124	3.3	609	8	O79569	O79569 mustelus ma
10	123.5	3.3	269	11	Q91VY8	Q91vy8 mus musculu
11	122	3.3	636	6	Q9GJX6	Q9gjx6 sus scrofa
12	122	3.3	643	6	Q9GJX7	Q9gjx7 sus scrofa
13	122	3.3	766	17	Q9UYP5	Q9uyp5 pyrococcus
14	121.5	3.3	591	8	Q36149	Q36149 trachemys s
15	121	3.3	605	8	Q9B6Y4	Q9b6y4 casuarius b
16	119	3.2	427	16	O53515	O53515 mycobacteri
17	119	3.2	612	8	Q9G6Q3	Q9g6q3 diplophos t
18	118.5	3.2	449	2	Q9RL05	Q9rl05 streptomyce
19	118.5	3.2	605	8	Q957Z0	Q957z0 casuarius c
20	118.5	3.2	613	8	Q94SY9	Q94sy9 crenimugil
21	118.5	3.2	694	2	Q9L179	Q9l179 streptomyce
22	117.5	3.2	583	5	O17069	O17069 caenorhabdi
23	117	3.1	604	8	Q9B6Z4	Q9b6z4 apteryx man
24	116.5	3.1	570	8	Q9B510	Q9b510 tetrodontop
25	116.5	3.1	613	8	Q94SZ9	Q94sz9 mugil cepha
26	116	3.1	605	8	Q9B6W0	Q9b6w0 dinornis gi
27	116	3.1	605	8	Q9B6U8	Q9b6u8 dromaius no
28	116	3.1	605	8	Q958A0	Q958a0 dromaius no
29	115.5	3.1	590	17	Q96XC0	Q96xc0 sulfolobus
30	115	3.1	317	16	Q97K74	Q97k74 clostridium
31	115	3.1	603	8	Q9MR50	Q9mr50 ciconia cic
32	115	3.1	603	8	Q958D4	Q958d4 tinamus maj
33	114.5	3.1	603	8	Q9GA20	Q9ga20 tupaia bela
34	114.5	3.1	829	10	Q9LMJ1	Q9lmj1 arabidopsis
35	114	3.1	395	2	Q9Z404	Q9z404 pseudomonas
36	114	3.1	606	8	Q9TA19	Q9ta19 loxodonta a
37	114	3.1	1201	16	Q9HU70	Q9hu70 pseudomonas
38	113.5	3.1	276	2	Q9ZBN3	Q9zbn3 streptomyce
39	113.5	3.1	605	8	Q957X8	Q957x8 anomalopter
40	113	3.0	604	8	Q9B6X2	Q9b6x2 crypturellu
41	113	3.0	689	13	Q9W701	Q9w701 xenopus lae
42	112	3.0	438	10	Q9ZU86	Q9zu86 arabidopsis
43	111.5	3.0	562	16	Q92T98	Q92t98 rhizobium m
44	111	3.0	544	17	Q9HQN5	Q9hqn5 halobacteri
45	111	3.0	614	8	Q94YP6	Q94yp6 osteoglossu

ALIGNMENTS

RESULT 1
Q9H720
ID Q9H720 PRELIMINARY; PRT; 699 AA.
AC Q9H720;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE CDNA: FLJ21511 FIS, CLONE COL05748.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=COLON;
RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
RA Okitani R., Ota T., Suzuki Y., Obayashi M., Nishi T., Shibahara T.,
RA Tanaka T., Nakamura Y., Isogai T., Sugano S.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AK025164; BAB15080.1; -.
SQ SEQUENCE 699 AA; 78565 MW; 2F839563189A0523 CRC64;

Query Match 99.8%; Score 3713; DB 4; Length 699;
Best Local Similarity 99.9%; Pred. No. 2.7e-266;
Matches 698; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 MTSLWREILLESLLGCVSWSLYHDLGPMIYYFPLQTLELTGLEGFSIAFLSPIFLTITPF 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MTSLWREILLESLLGCVSWSLYHDLGPMIYYFPLQTLELTGLEGFSIAFLSPIFLTITPF 60

Qy      61 WKLVNKKWMLTLLRIITIGSIASFQAPNAKLRLMVLALGVSSSLIVQAVTWWSGSHLQRY 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 WKLVNKKWMLTLLRIITIGSIASFQAPNAKLRLMVLALGVSSSLIVQAVTWWSGSHLQRY 120

Qy     121 LRIWGFILGQIVLVVLRIWYTSLNPIWSYQMSNKVILTLSAIATLDRIGTDGDCSKPEEK 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 LRIWGFILGQIVLVVLRIWYTSLNPIWSYQMSNKVILTLSAIATLDRIGTDGDCSKPEEK 180
```

```
Qy    181 KTGEVATGMASRPNWLLAGAAFGSLVFLTHWVFGEVSLVSRWAVSGHPHPGPDPNPFGGA 240
Db    181 KTGEVATGMASRPNWLLAGAAFGSLVFLTHWVFGEVSLVSRWAVSGHPHPGPDPNPFGGA 240

Qy    241 VLLCLASGLMLPSCLWFRGTGLIWWVTGTASAAGLLYLHTWAAAVSGCVFAIFTASMWPQ 300
Db    241 VLLCLASGLMLPSCLWFRGTGLIWWVTGTASAAGLLYLHTWAAAVSGCVFAIFTASMWPQ 300

Qy    301 TLGHLINSGTNPGKTMTIAMIFYLLEIFFCAWCTAFKFVPGGVYARERSDVLLGTMMLII 360
Db    301 TLGHLINSGTNPGKTMTIAMIFYLLEIFFCAWCTAFKFVPGGVYARERSDVLLGTMMLII 360

Qy    361 GLNMLFGPKKNLDLLLQTKNSSKVLFRKSEKYMKLFLWLLVGVGLLGLGLRHKAYERKLG 420
Db    361 GLNMLFGPKKNLDLLLQTKNSSKVLFRKSEKYMKLFLWLLVGVGLLGLGLRHKAYERKLG 420

Qy    421 KVAPTKEVSAAIWPFRFGYDNEGWSSLERSAHLLNETGADFITILESDASKPYMGNNDLT 480
Db    421 KVAPTKEVSAAIWPFRFGYDNEGWSSLERSAHLLNETGADFITILESDASKPYMGNNDLT 480

Qy    481 MWLGEKLGFYTDFGPSTRYHTWGIMALSRYPIVKSEHHLLPSPEGEIAPAITLTVNISGK 540
Db    481 MWLGEKLGFYTDFGPSTRYHTWGIMALSRYPIVKSEHHLLPSPEGEIAPAITLTVNISGK 540

Qy    541 LVDFVVTHFGNHEDDLDRKLQAIAVSKLLKSSSNQVIFLGYITSAPGSRDYLQLTEHGNV 600
Db    541 LVDFVVTHFGNHEDDLDRKLQAIAVSKLLKSSSNQVIFLGYITSAPGSRDYLQLTEHGNV 600

Qy    601 KDIDSTDHDRWCEYIMYRGLIRLGYARISHAELSDSEIQMAKFRIPDDPTNYRDNQKVVI 660
Db    601 KDIDSTDHDRWCEYIMYRGLIRLGYARISHAELSDSEIQMAKFRIPDDPTNYRDNQKVVI 660

Qy    661 DHREVSEKIHFNPRFGSYKEGHNYENNHHFHMNTPKYFL 699
Db    661 DHREVSEKIHFNPRFGSYKEGHNYENNHNFHMNTPKYFL 699
```

RESULT 2
Q91YL7
ID Q91YL7 PRELIMINARY; PRT; 699 AA.
AC Q91YL7;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE SIMILAR TO HYPOTHETICAL PROTEIN FLJ21511.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (OCT-2001) to the EMBL/GenBank/DDBJ databases.
DR EMBL; BC016523; AAH16523.1; -.
KW Hypothetical protein.
SQ SEQUENCE 699 AA; 78179 MW; 4F41975D6570D5F8 CRC64;

Query Match 84.0%; Score 3125; DB 11; Length 699;
Best Local Similarity 82.8%; Pred. No. 8.5e-223;
Matches 579; Conservative 41; Mismatches 79; Indels 0; Gaps 0;

```
Qy      1 MTSLWREILLESLLGCVSWSLYHDLGPMIYYFPLQTLELTGLEGFSIAFLSPIFLTITPF 60
Db      1 MPGLWRAIALETLLGYVSWSLYHGLSPMIYYFPLQTLELTGLEFFCVAFLSPILLTIPPL 60

Qy     61 WKLVNKKWMLTLLRIITIGSIASFQAPNAKLRLMVLALGVSSSLIVQAVTWWSGSHLQRY 120
Db     61 WKLVNKKWTLSLLRIVTVGSIASFEAPNAKLRLMVLALGVSSSLIVQTVTWWSGSGLQRY 120

Qy    121 LRIWGFILGQIVLVVLRIWYTSLNPIWSYQMSNKVILTLSAIATLDRIGTDGDCSKPEEK 180
Db    121 LKIWGFILGHVLLLVLRIWYTSLNPIWSYQMSNRVILTLSAVAVLDRIGTDGDYRNPEGK 180
```

```
Qy    181 KTGEVATGMASRPNWLLAGAAFGSLVFLTHWVFGEVSLVSRWAVSGHPHPGPDPNPFGGA 240
Db    181 KPREVATGRTSLSSWLLPGAAFGSLLFLTHWIFGEVSIVSRWAVSGHPHPGPDPNPFGGA 240

Qy    241 VLLCLASGLMLPSCLWFRGTGLIWWVTGTASAAGLLYLHTWAAAVSGCVFAIFTASMWPQ 300
Db    241 VLLGFSSGLMLSGSSWLHDAGLAWWMTGAASAMGLLYLRTWAAAVSGCVLAVFTGSMWPQ 300

Qy    301 TLGHLINSGTNPGKTMTIAMIFYLLEIFFCAWCTAFKFVPGGVYARERSDVLLGTMMLII 360
Db    301 VLGHLVNSGKNSGEAMATGMILYVLQTFFCAWCTAFKFVPGGVYARERSDVLLGTIMVII 360

Qy    361 GLNMLFGPKKNLDLLLQTKNSSKVLFRKSEKYMKLFLWLLVGVGLLGLGLRHKAYERKLG 420
Db    361 GLSMLFGPKKNLDFLLQTKNSPKTLLRCSEKYMKLILWLFVGVGLLGLGLRHRTYERQLG 420

Qy    421 KVAPTKEVSAAIWPFRFGYDNEGWSSLERSAHLLNETGADFITILESDASKPYMGNNDLT 480
Db    421 RGAPATVVSAAIWPFRFGYDNEGWPNLERSAQLLKETGADFITILESDASKPYIGNNDLT 480

Qy    481 MWLGEKLGFYTDFGPSTRYHTWGIMALSRYPIVKSEHHLLPSPEGEIAPAITLTVNISGK 540
Db    481 MWLGEKLGFYTDFGPSTRDHTWGIMVLSRYPIVRSEHHLLPSPEGEIAPAITMTVNVSNR 540

Qy    541 LVDFVVTHFGNHEDDLDRKLQAIAVSKLLKSSSNQVIFLGYITSAPGSRDYLQLTEHGNV 600
Db    541 LVDFVVTHFGNHEDDLDRKLQAIAVSKLLKNCSNQVIFLGYITSEPGSRDYIQLTKHGNV 600

Qy    601 KDIDSTDHDRWCEYIMYRGLIRLGYARISHAELSDSEIQMAKFRIPDDPTNYRDNQKVVI 660
Db    601 KDIDSSDGDRWCEYIMYRGLIRLGYARISHAELSDSEIQMAKFRIPDDPANYRDNQKVVI 660

Qy    661 DHREVSEKIHFNPRFGSYKEGHNYENNHHFHMNTPKYFL 699
Db    661 DHRGVPKNIHFNPRFGSYKEGHNYENTHHFHMNTPKYFV 699
```

RESULT 3
Q9P6M4
ID Q9P6M4 PRELIMINARY; PRT; 425 AA.
AC Q9P6M4;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE HYPOTHETICAL 48.6 KDA PROTEIN (FRAGMENT).
GN SPAC688.01.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomyces.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972H-;
RA Aert R., Robben J., Volckaert G., Wood V., Rajandream M.A.,
RA Barrell B.G.;
RL Submitted (APR-2000) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AL355632; CAB90768.1; -.
KW Hypothetical protein.
FT NON_TER 1 1
SQ SEQUENCE 425 AA; 48575 MW; 1B17FE034278A0E4 CRC64;

Query Match 17.2%; Score 639; DB 3; Length 425;
Best Local Similarity 34.2%; Pred. No. 3.6e-39;
Matches 152; Conservative 70; Mismatches 178; Indels 44; Gaps 10;

```
Qy    276 LYL-HTWAAAVSGCVFAIFT-----ASMWPQTLGHLINSGTNPGKTMTIAMIFYLLEIFF 329
Db      4 MYINHGWCSYLGGLIFTSYVLIYSFASI-------RISSFYSPAKVWGGAFLVYILYSLA 56

Qy    330 CAWCTAFKFVPGGVYARERSDVLLGTMMLIIGLNMLF----------GPKKNLDLLLQTK 379
Db     57 HVWVVAYEFVPGGPILRERTSYIL----IFIGWNLAALVPAYSGESKEPNKADSSVVDIK 112
```

```
Qy    380 NSSKVLFRKSEKYMKLFLWLLVGVGLLGLGLRHKAYERKLGKVAP----TKEVSAAIWPF 435
          |     |:| |       || |  |  : |:            |     |  :| ||
Db    113 QSDSSYRRRSFKKS-----LLTGFCLALMALKFAIQNMPPYDYTPYHPNEKLFTAGIWTI 167

Qy    436 RFGYDNEGWSSLERSAHLLNETGADFITILESDASKPYMGNNDLTMWLGEKLGFYTDFGP 495
          || ||  ::|  |    : :   |   :|||| :  ||  |||  |   || | |:||
Db    168 HFGLDNFMYASENRIRDAVRDMELDVFGLLESDTQRLIMGFRDLTQVLAHDLGMYADYGP 227

Qy    496 STRYHTWGIMALSRYPIVKSEHHLLPSPEGEIAPAITLTVNISGKLVDFVVTHFGNHEDD 555
            ||||   ||::||| | |||||||:||:||||  |::: |:|:| ||:| | :|
Db    228 GPDKHTWGAALLSKFPIVNSTHHLLPSPQGELAPAIHATLDVYGELIDVVVSHNGQYESQ 287

Qy    556 LDRKLQAIAVSKLLKSSSNQVIFLGYITSAPGSRDYLQLTEHGNVKDIDSTDHDRWCEYI 615
          |||:||:  :::::: |   ::|||: |  |      ||    : ||:  |:||||:||
Db    288 LDRRLQSTELARIMRESPRPLVFLGYVVSNVGQEPQTILTRDTGMLDIEPADYDRWCQYI 347

Qy    616 MYRGLIRLGYARISHAELSDSEIQMAKFRIPDDPTNYRDNQKVVIDHREVSEKIHFNPRF 675
          |||: |:||||:  : ::|:|:|  || :  |      : | ||   | |   :   |
Db    348 FYRGVKRIGYARLHRSTITDTELQTGKFLVTKDL-----GRNVRIDKEHVPESHRYPSLF 402

Qy    676 -GSYKEGHNYENNHHFHMNTPKYF 698
          |:   || |:||   |   | |:
Db    403 EGTGVNGHYYDNNLVVH--EPWYY 424


RESULT 4
P78781
ID   P78781      PRELIMINARY;      PRT;   401 AA.
AC   P78781;
DT   01-MAY-1997 (TrEMBLrel. 03, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   UNKNOWN PROTEIN (FRAGMENT).
OS   Schizosaccharomyces pombe (Fission yeast).
OC   Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC   Schizosaccharomycetales; Schizosaccharomycetaceae;
OC   Schizosaccharomyces.
OX   NCBI_TaxID=4896;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=PR745;
RX   MEDLINE=98162722; PubMed=9501991;
RA   Yoshioka S., Kato K., Nakai K., Okayama H., Nojima H.;
RT   "Identification of open reading frames in Schizosaccharomyces pombe
RT   cDNAs.";
RL   DNA Res. 4:363-369(1997).
DR   EMBL; D89130; BAA13792.1; -.
FT   NON_TER       1      1
SQ   SEQUENCE   401 AA;  45758 MW;  53817CE5A310F2D8 CRC64;

  Query Match             16.8%;  Score 625.5;  DB 3;  Length 401;
  Best Local Similarity   35.5%;  Pred. No. 3.3e-38;
  Matches  145;  Conservative   63;  Mismatches  169;  Indels   31;  Gaps    7;

Qy    306 INSGTNPGKTMTIAMIFYLLEIFFCAWCTAFKFVPGGVYARERSDVLLGTMMLIIGLNML 365
          |:|  :| |    | : |:|      |  |::|||||   |||:  :|    : || |:
Db      9 ISSFYSPAKVWGGAFLVYILYSLAHVWVVAYEFVPGGPILRERTSYIL----IFIGWNLA 64

Qy    366 F----------GPKKNLDLLLQTKNSSKVLFRKSEKYMKLFLWLLVGVGLLGLGLRHKAY 415
                     | |    ::  | |     |:| |       || |  |  : |:
Db     65 ALVPAYSGESKEPNKADSSVVDIKQSDSSYRRRSFKKS-----LLTGFCLALMALKFAIQ 119

Qy    416 ERKLGKVAP----TKEVSAAIWPFRFGYDNEGWSSLERSAHLLNETGADFITILESDASK 471
                |     |  :| ||   || ||  ::|  |    : :   |   :|||| :
Db    120 NMPPYDYTPYHPNEKLFTAGIWTIHFGLDNFMYASENRIRDAVRDMELDVFGLLESDTQR 179

Qy    472 PYMGNNDLTMWLGEKLGFYTDFGPSTRYHTWGIMALSRYPIVKSEHHLLPSPEGEIAPAI 531
            ||  |||  |   || | |:||    ||||   ||::||| | |||||||:||:||||
Db    180 LIMGFRDLTQVLAHDLGMYADYGPGPDKHTWGAALLSKFPIVNSTHHLLPSPQGELAPAI 239

Qy    532 TLTVNISGKLVDFVVTHFGNHEDDLDRKLQAIAVSKLLKSSSNQVIFLGYITSAPGSRDY 591
          |::: |:|:| ||:| | :|  |||:||:  :::::: |   ::||||: |  |

Db    240 HATLDVYGELIDVVVSHNGQYESQLDRRLQSTELARIMRESPRPLVFLGYVVSNVGQEPQ 299

Qy    592 LQLTEHGNVKDIDSTDHDRWCEYIMYRGLIRLGYARISHAELSDSEIQMAKFRIPDDPTN 651
            ||    : ||:  |:||||:|| |||: |:||||:  : ::|:|:|  || :  |
Db    300 TILTRDTGMLDIEPADYDRWCQYIFYRGVKRIGYARLHRSTITDTELQTGKFLVTKDL-- 357

Qy    652 YRDNQKVVIDHREVSEKIHFNPRF-GSYKEGHNYENNHHFHMNTPKYF 698
              : | ||   | |   :   | |:   || |:||   |   | |:
Db    358 ---GRNVRIDKEHVPESHRYPSLFEGTGVNGHYYDNNLVVH--EPWYY 400


RESULT 5
Q9HDZ2
ID   Q9HDZ2      PRELIMINARY;      PRT;   688 AA.
AC   Q9HDZ2;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   HYPOTHETICAL 77.5 KDA PROTEIN (FRAGMENT).
GN   SPAC589.12.
OS   Schizosaccharomyces pombe (Fission yeast).
OC   Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC   Schizosaccharomycetales; Schizosaccharomycetaceae;
OC   Schizosaccharomyces.
OX   NCBI_TaxID=4896;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972H-;
RA   Wood V., Rajandream M.A., Barrell B.G., Aert R., Robben J.,
RA   Weltjens I., Grymonprez B., Volckaert G.;
RL   Submitted (JAN-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL512496; CAC19769.1; -.
KW   Hypothetical protein.
FT   NON_TER      688    688
SQ   SEQUENCE   688 AA;  77474 MW;  3985EAC23CDA34FF CRC64;

  Query Match             6.9%;  Score 256;  DB 3;  Length 688;
  Best Local Similarity   22.5%;  Pred. No. 1.5e-10;
  Matches   97;  Conservative   77;  Mismatches  197;  Indels   60;  Gaps   11;

Qy     11 ESLLGCVSWSLYHDLGPMIYYFPLQTLELTGLEGFSIAFLSPIFLTITPFWKLVNKKWML 70
          |  |  | ||:   || :::|||| : ::| |   :  |||  | |    | :|  ::
Db    285 EVYLSFVFWSVLTSLGLLVWYFPLWHMGISGYEACILFELSPFLLGIPLLRKFASKVPVI 344

Qy     71 TLLRIITIGSIASFQAPNAKLRLMVLALGVSSSLIVQAVTWW------SGSHLQRYLRIW 124
            |   : :  ||:::  :   || | |     |:  : : |       |  :|    :|
Db    345 FL--FLNVIGIAAYKLEDPVHRLFVTAF----SVCCECLAWTSLFSNISPENLAIERKIS 398

Qy    125 GFILGQIVLVVLRIWYTSLNPIW--------SYQMSNKVILTLSAIATLDRIGTDGDCSK 176
          |: | :   : :  : | ||||          |:   ::  :: :              :
Db    399 TFLFGLLASSIAKYSFFSNNPIWPILNETNGGKQIPALIVGIIACLIFAIFHVQQTTANA 458

Qy    177 PEEKKTGEVATGMASRPNWLLAGAAFGSLVFLTHWVFGEVSLVSRWAVSGHPHPGPDPNP 236
          |  |  ::          | |  : |:::|  |    : :::  |:  |:|  || | |
Db    459 VEHFKLRKITA--------LSAALSLGTVLFCLHTFLCDSTVLMTWSWDGYPIKGPQPYP 510

Qy    237 FGGAVLLCLASGLMLPSCLWFRGTGLIWWVTGTASAAGLLYL-HTWAAAVSGCVFAIFT- 294
          |   ::      :::   |:  |  ::          :  :|: | | : : | :|  :
Db    511 HGAVSIVVSICAVLVAPYLYQSGAFMLIGFVLACFGSYFMYINHGWCSYLGGLIFTSYVL 570

Qy    295 ----ASMWPQTLGHLINSGTNPGKTMTIAMIFYLLEIFFCAWCTAFKFVPGGVYARERSD 350
              ||:        |:|  :| |    | : |:|       |  |::|||||   |||:
Db    571 IYSFASI-------RISSFYSPAKVWGGAFLVYILYSLAHVWVVAYEFVPGGPILRERTS 623

Qy    351 VLLGTMMLIIGLNMLF----------GPKKNLDLLLQTKNSSKVLFRKSEKYMKLFLWLL 400
            :|    : || |:              | |    ::  | |     |:| |       ||
Db    624 YIL----IFIGWNLAALVPAYSGESKEPNKADSSVVDIKQSDSSYRRRSFKKS-----LL 674

Qy    401 VGVGLLGLGLR 411
          |  |  : |:
Db    675 TGFCLALMALK 685
```

```
RESULT 6
Q9S1Z5
ID   Q9S1Z5      PRELIMINARY;      PRT;   279 AA.
AC   Q9S1Z5;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   PUTATIVE SECRETED PROTEIN.
GN   SCJ21.12C.
OS   Streptomyces coelicolor.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=1902;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RA   Seeger K.J., Harris D.;
RL   Submitted (AUG-1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RA   Parkhill J., Barrell B.G., Rajandream M.A.;
RL   Submitted (AUG-1999) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RX   MEDLINE=97000351; PubMed=8843436;
RA   Redenbach M., Kieser H.M., Denapaite D., Eichner A., Cullum J.,
RA   Kinashi H., Hopwood D.A.;
RT   "A set of ordered cosmids and a detailed genetic and physical map for
RT   the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL   Mol. Microbiol. 21:77-96(1996).
DR   EMBL; AL109747; CAB52358.1; -.
SQ   SEQUENCE   279 AA;  30031 MW;  B16719A1698580C0 CRC64;

  Query Match             3.8%;  Score 141.5;  DB 2;  Length 279;
  Best Local Similarity   24.9%;  Pred. No. 0.013;
  Matches   63;  Conservative   36;  Mismatches   91;  Indels   63;  Gaps   12;

Qy     447 LERSAHLLNETGADFITILESDASKPYMGNNDLTMW------LGEKLGFYTDFG------ 494
           | | |::: ::||| : : | |  | |   :|   |      | | ||::  ||
Db      58 LRRVANVIRKSGADVVGLQEVD--KHYSARSD---WADQPAELAELLGYHVVFGANIDNS 112

Qy     495 ---PSTRYHTWGIMALSRYPIVKSEH-HLLPSPEGEIAPAITLTVNISGKLVDFVVTHF- 549
               |     :|   |||||| |::  |  ||  |    :  |::: || |:|  ||
Db     113 PPAPGGHRVQYGTAILSRYPITASDNTWLYKSPGQEQRGLLHATLDVHGKKVEFYNTHLA 172

Qy     550 -GNHEDDLDRKLQAIAVSKLLKSSSNQVIFLGYITSAPGSRDYLQLTEHGNVKDIDSTDH 608
            |:  | | :  | : :    |      | :|   : | : :          : : :
Db     173 AGSQADRLQQTAQVVDLIGTRKPG----ILVGDFNALPAAPE---------SRPLQNAYT 219

Qy     609 DRWCEYIMYRG-----------LIRLGYARISHAELSDSEIQMAKFRIPDDPTNYRDNQ 656
           | | : :  ||          | | ||      ::    |: |    |||
Db     220 DAWAKSLHARGDGATYPAQSPTERIDLIYAT---RRVTPLVAQVLK----DDP------- 265

Qy     657 KVVIDHREVSEKI 669
               |||  :  |:
Db     266 -AASDHRPLLGKV 277


RESULT 7
Q950T2
ID   Q950T2      PRELIMINARY;      PRT;   583 AA.
AC   Q950T2;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   NADH DEHYDROGENASE SUBUNIT 5 (EC 1.6.5.3).
GN   NAD5.
OS   Hyaloraphidium curvatum.
OG   Mitochondrion.
OC   Eukaryota; Fungi; Fungi incertae sedis; Hyaloraphidium.
OX   NCBI_TaxID=82268;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Forget L., Ustinova J., Wang Z., Huss V.A.R., Lang F.B.F.;
RT   "Hyaloraphidium curvatum: a linear mitochondrial genome, tRNA editing,
RT   and an evolutionary link to lower fungi.";
RL   Mol. Biol. Evol. 0:0-0(2001).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Lang F.B.F.;
RL   Submitted (JUL-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF402142; AAK83434.1; -.
KW   Oxidoreductase; Mitochondrion.
SQ   SEQUENCE   583 AA;  64424 MW;  2B568D4847987022 CRC64;

  Query Match             3.5%;  Score 132;  DB 8;  Length 583;
  Best Local Similarity   19.4%;  Pred. No. 0.18;
  Matches  116;  Conservative   91;  Mismatches  194;  Indels  196;  Gaps   28;

Qy       8 ILLESLLGCVSWSLYHDLGPMIYYF--PLQTLELTGLEGFSIAFLSPIFLTITPFWKLVN 65
           |:   ||| :|  |   ||  : |   || |: | |:          : | :  :::::
Db       5 IITLPLLGSLSSGL---LGRWLGYSGGPLLTVLLMGV---------TLALVLCGYYEIIF 52

Qy      66 KKWMLTL---LRIITIGSI---ASFQAPNAKLRLMVLALGVSSSLIVQAVTWWSGS-HLQ 118
             :  | |   :  :|:| :    ||       : :|:    ||  : : |: :  |  |:|
Db      53 HQSSLLLPLPISWLTLGWLDLDMSFIMDELSISMMIPICTVSFLVHMYAIGYMHGDPHIQ 112

Qy     119 R---YLRIWGFIL------------------GQIVLVVLRIWYTSLNPIWSYQMSNKVI 156
           |   || :: | :                  | :  :::  |:| |    : | : :
Db     113 RFFSYLSLFTFFMLLMVTADNWLLLFIGWEGVGLVSYLLIGFWFTRLR---AGQAALQAF 169

Qy     157 LTLSAIATLDRIGTDGDCSKPEEKKTGEVATGMASRPNWLLAGAAFGSLVFLTHWVFGEV 216
           |       ::|||  |                  || |       :  |: |::
Db     170 L-------MNRIGDTG---------------------LFLAMA-------IAIWLTGDL 193

Qy     217 SLVSRWAVSGHPHPGPDPNPFGGAVLLCLASGLMLPSCLWFRGTGLIWWVTGTASAAGLL 276
             : :|:  | :|                          |    ||:  :  ||  :| |
Db     194 EFSTFFALLPHMNP----------------------YWTALIGLLMLLAVTAK-SGQL 228

Qy     277 YLHTW-AAAVSGCVFAIFTASMWPQTLGHLINSGTNPGKTMTIAMIFYLLE---IF---- 328
           || |   |: |           |  :  ||::      ||  | |: ||    :|
Db     229 GLHAWLPVAMEG----------PTPVSALIHAA-----TMVTAGIYLLLRFNALFGFSP 272

Qy     329 FCAWCTAFKFVPGGVYARERSDV----------LLGTMMLIIGLNMLFGPKKNLDLLLQT 378
             |  |   : | ||   ::|:           || |:|  ||
Db     273 ILIWLGAATALFGAVYGYVQTDLKRTIAYSTTSQLGYMVLACGLG---------------- 317

Qy     379 KNSSKVLFRKSEKYMKLFLWLLVGVGLLGLGLRHKAYE----RKLGKVAPTKEV------ 428
           :    :    :  : |  |:|  |       : |  ::    ||:| :     |
Db     318 QYGLALAHLVNHAFFKALLFLSAG------SVIHAIHDEQDIRKMGGIMSMMPVTYITSM 371

Qy     429 ----SAAIWPFRFGYDNEGW---SSLERSAHLLNETGADFITILES----------DAS 470
               |    ||  || :: :   ::   : :|:  || : |   |             |
Db     372 VGSLSLVALPFLTGYYSKDFILQTAFGTTGYLMG-LGAAYFTAFYSLKLLHRVFWLPPQS 430

Qy     471 KPYMGNNDLTMWLGEKLGFYTDFGPSTRYHTWGIMALSRYPIVKSEHHLLPSPEGEI 527
           |  :| :: : |:       | |  |     || :| :   :  : : |   |: :|
Db     431 KVVLGAHEPSAWMLIPTTTLTLFSIS-----WGYLAQNHAALAPAMNTLFLLPQNQI 482


RESULT 8
P95556
ID   P95556      PRELIMINARY;      PRT;   396 AA.
AC   P95556;
DT   01-MAY-1997 (TrEMBLrel. 03, Created)
DT   01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   ORF396 PROTEIN.
GN   ORF396.
OS   Pseudomonas stutzeri (Pseudomonas perfectomarina).
OC   Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC   Pseudomonas.
OX   NCBI_TaxID=316;
RN   [1]
```

```
RP   SEQUENCE FROM N.A.
RC   STRAIN=ZOBELL ATCC 14405;
RX   MEDLINE=97107629; PubMed=8950369;
RA   Glockner A.B., Zumft W.G.;
RT   "Sequence and analysis of an internal 9.7-kb segment from the 30-kb
RT   denitrification gene cluster of Pseudomonas stutzeri.";
RL   Biochim. Biophys. Acta 1277:6-12(1996).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ZOBELL ATCC 14405;
RX   MEDLINE=96028114; PubMed=7588711;
RA   Palmedo G., Seither P., Koerner H., Matthews J.C., Burkhalter R.S.,
RA   Timkovich R., Zumft W.G.;
RT   "Resolution of the nirD locus for heme d1 synthesis of cytochrome cd1
RT   (respiratory nitrite reductase) from Pseudomonas stutzeri.";
RL   Eur. J. Biochem. 232:737-746(1995).
DR   EMBL; Z73914; CAA98157.1; -.
SQ   SEQUENCE   396 AA;  43083 MW;  EB07255ECFB1E0D3 CRC64;

  Query Match             3.5%;  Score 128.5;  DB 2;  Length 396;
  Best Local Similarity   21.1%;  Pred. No. 0.19;
  Matches   82;  Conservative   55;  Mismatches  147;  Indels  105;  Gaps   20;

Qy      55 LTITPFWKLVNKKWML--TLLRIITI--------GSIASFQAPNAKL------RLMVLAL 98
           |:| | |:|  : : |  :|  :: |         |    ||      |        |   |:
Db      10 LSIAPIWRLAFRPFFLAGSLYALLAIPLWVAAWTGLWPGFQPTGGWLAWHRHEMLFGFAM 69

Qy      99 GVSSSLIVQAVTWW------SGSHLQRYLRIW-----GFILG---------QIVLVVLRI 138
             : :  :: ||  |      ||: |     :|     |:: |           :: :|  :
Db      70 AIVAGFLLTAVQTWTGQTAPSGNRLVGLAAVWLAARLGWLFGLPAAWLAPLDLLFLVALV 129

Qy     139 WYTSLNPIWSYQMSNK--VILTLSAIATLDRIGTDGDCSKPEEKKTGEVATGMASRPNWL 196
           |  :   :|: :      ::: || :   | :   |     :  :   |  |:     ||
Db     130 WMMA-QMLWAVRQKRNYPIVVVLSLMLGADVLILTGLLQGNDALQRQGVLAGL-----WL 183

Qy     197 LAG--AAFGSLV--FLTH----------WVFGEVSLVSRWAVSGHPHP---GPDPNPFG 238
           :|   |  |  |  | |            ||: :|:|:    |    |       | |
Db     184 VAALMALIGGRVIPFFTQRGLGKVDAVKPWVWLDVALLVGTGVIALLHAFGVAMRPQPLL 243

Qy     239 GAVLLCLASGLMLPSCLWF-RG---TGLIW-------WVTGTASAAGLLYLHTWAAAVS- 286
           | : : :  | :|    |: :|    ||:|       |:    :| ||   |    | |
Db     244 GLLFVAIGVGHLLRLMRWYDKGIWKVGLLWSLHVAMLWL--VVAAFGLALWHFGLLAQSS 301

Qy     287 --------GCVFAIFTASMWPQTLGH---------------LINSGTNPGKTMTIAMIF 322
                 | :  :  | :  ||||                 | | ||     :::|
Db     302 PSLHALSVGSMSGLILAMIARVTLGHTGRPLQLPAGIIGAFVLFNLGTAARVFLSVAWPV 361

Qy     323 YLLEIFFCAWCTAF-----KFVPGGVYAR 346
             | :    |  ||     :: |  | ||
Db     362 GGLWLAAVCWTLAFALYVWRYAPMLVAAR 390


RESULT 9
O79569
ID   O79569      PRELIMINARY;      PRT;   609 AA.
AC   O79569;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   NADH DEHYDROGENASE SUBUNIT 5.
GN   NADH5.
OS   Mustelus manazo.
OG   Mitochondrion.
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;
OC   Elasmobranchii; Galeomorphii; Galeoidea; Carcharhiniformes; Triakidae;
OC   Mustelus.
OX   NCBI_TaxID=79736;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=LIVER;
RX   MEDLINE=99083431; PubMed=9866199;
RA   Cao Y., Waddell P.J., Okada N., Hasegawa M.;
RT   "The complete mitochondrial DNA sequence of the shark (Mustelus
RT   manazo): Evaluating rooting contradictions to living bony
RT   vertebrates.";
RL   Mol. Biol. Evol. 15:1637-1646(1998).
CC   -!- CATALYTIC ACTIVITY: NADH + UBIQUINONE = NAD(+) + UBIQUINOL.
DR   EMBL; AB015962; BAA33045.1; -.
DR   InterPro; IPR003916; NADHub_oxdrdctse5.
DR   InterPro; IPR001750; Oxidored_q1.
DR   InterPro; IPR001516; Oxidored_q1_N.
DR   InterPro; IPR000215; Serpin.
DR   Pfam; PF00361; oxidored_q1; 1.
DR   Pfam; PF00662; oxidored_q1_N; 1.
DR   PRINTS; PR01434; NADHDHGNASE5.
DR   PROSITE; PS00284; SERPIN; UNKNOWN_1.
KW   Mitochondrion; NAD; Oxidoreductase; Ubiquinone.
SQ   SEQUENCE   609 AA;  68215 MW;  97FBE6D34CDE7DD0 CRC64;

  Query Match             3.3%;  Score 124;  DB 8;  Length 609;
  Best Local Similarity   19.3%;  Pred. No. 0.74;
  Matches  131;  Conservative   83;  Mismatches  210;  Indels  254;  Gaps   32;

Qy     155 VILTLSAIATLDRIGTDGDCSKPEEKKTGEVATGMASRPNW--LLAGAA-----FGSLV- 206
           :|||| |::|           |:| |           |||  |    |     | ||:
Db      14 IILTLPLISSL----------SPKELK----------PNWSSLYVKTAVKISFFISLIP 52

Qy     207 ----------FLTHW--------------------VFGEVSLVSRWAV----SGHPHP 230
                    :|:|                    :|  |:|   |::      : |
Db      53 LFIFLDQGLESIVTNWNWMNMGPFDINMSFKFDLYSIIFTPVALYVTWSILEFALWYMHS 112

Qy     231 GPDPNPFGGAVLLCLASGLMLPSC-------LWFRGTGLI------WWVT----GTASAA 273
           |: | |   :|| | | ::| :         : : | |::      || :     ||:
Db     113 DPNMNRFFKYLLLFLISMIILVTANNMFQLFIGWEGVGIMSFLLIGWWYSRADANTAALQ 172

Qy     274 GLLY-----------------LHTW----------------------AAAVSGCVFAI 292
           ::|                  |::|                      |||     | :
Db     173 AVIYNRIGDVGLILSMAWLATNLNSWEIHQLFILSKNKDLTLPLLGLVLAAAGKSAQFGL 232

Qy     293 F----TASMWPQTLGHLINSGTNPGKTMTIAMIFYLLEIF----------FCAWCTAFK 337
           :|   |  :  |::|      || :| || |: :             |   |
Db     233 HPWLPSAMEGPTPVSALLHS-----STMVVAGIFLLIRLHPLIQDNKLILTVCLCLGALT 287

Qy     338 FVPGGVYARERSDV----------LLGTMMLIIGLNMLFGPKKNLDLLLQTKNSSKVLFR 387
           :      |  ::|:           || ||: ||||                :     |
Db     288 TLFTATCALTQNDIKKIVAFSTSSQLGLMMVTIGLN--------------QPQLAFLHI 332

Qy     388 KSEKYMKLFLWLLVGVGLLGLGLRHKAYE----RKLG---KVAPTKEVSAAIWPFRFGYD 440
           :  : |  |:|  |       : |   :    ||:|   |: |    |  |
Db     333 CTHAFFKAMLFLCSG------SIIHSLNDEQDIRKMGGLHKLLPFTSTSLTIGS------ 380

Qy     441 NEGWSSLERSAHLLNETGADFIT-ILESDASKPYMGNNDLTMWLGEKLGFYTDFGPSTRY 499
                     |  ||  |::     ||     |  : |  |         | |   |
Db     381 -------------LALTGMPFLSGFFSKDAIIESMNTSHLNAWALILTLVATSF---TAI 424

Qy     500 HTWGIM--ALSRYP------IVKSEHHLLPSP-----EGEIAPAITLTVNISGKLVDFVV 546
           ::  ::  ||  ||       :   : |: :|      | |   : :|:|::
Db     425 YSLRLIFFALMNYPRFNTLSPINENNPLVINPIKRLAYGSIIAGLIITLNLT-------- 476

Qy     547 THFGNHEDDLDRKLQAIAVSKLLKSSSNQVIFLGYITSAPGSRDYLQLTEHGNVKDIDST 606
                   | | : :| ||| |:  |  :| : :       |:||   |
Db     477 ----------PTKTQIMTMSPLLKLSALLVTIMGLLLA-------LELTNLTN------- 512

Qy     607 DHDRWCEYIMYRGLIR-LGY-ARISHAELSDSEIQMAKF---RIPDDPTNYRDNQKVVID 661
           | :    : |      |||  | |  |  : :  |::    : |   | :   |  :
Db     513 SHFKTNPTLHYHHFSNLLGYFPSIIHRLLPKTSLNWAQYISTHLIDQTWNEKIGPKSNLI 572

Qy     662 HREVSEKIHFNPRFGSYK 679
           :    |:   |: |  |
Db     573 QQTSLIKLSTQPQQGLIK 590


RESULT 10
Q91VY8
ID   Q91VY8      PRELIMINARY;      PRT;   269 AA.
AC   Q91VY8;
```

```
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   AQUAPORIN 1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BREAST TUMOR;
RA   Strausberg R.;
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC007125; AAH07125.1; -.
KW   Porin.
SQ   SEQUENCE   269 AA;  28823 MW;  80534DE4B78AB5E7 CRC64;

  Query Match             3.3%;  Score 123.5;  DB 11;  Length 269;
  Best Local Similarity   22.8%;  Pred. No. 0.27;
  Matches   60;  Conservative   36;  Mismatches   82;  Indels   85;  Gaps   11;

Qy      60 FWKLVNKKWM-LTLLRIITIGSIASFQAPNAKLRLMV-------LALGVSSSLIVQAVTW 111
           ||: |  ::: :||   |:|||   |  |  : : :|       || |:| : : |:|
Db      10 FWRAVVAEFLAMTLFVFISIGSALGFNYPLERNQTLVQDNVKVSLAFGLSIATLAQSVGH 69

Qy     112 WSGSHLQRYLRIWGFILGQIVLVVLRIWY-----------TSLNPIWSYQMSNK----- 154
            ||:||   : : | :|   : ::  : |              |: | |  : |
Db      70 ISGAHLNPAVTL-GLLLSCQISILRAVMYIIAQCVGAIVATAILSGITSSLVDNSLGRND 128

Qy     155 -----------------VILTLSAIATLDRIGTDGDCSKPEEKKTGEVATGMASRPNWL 196
                            : | |  :|| ||   |   | |       :| |::     |
Db     129 LAHGVNSGQGLGIEIIGTLQLVLCVLATTDRRRRDLGGSAP-------LAIGLSVALGHL 181

Qy     197 LA----------GAAFGSLV----FLTHWVFGEVSLVSRWAVSGHPHPGPDPNPF-GGAV 241
           ||          :||| |    |  ||:|        |             || |||:
Db     182 LAIDYTGCSINPARSFGSAVLTRNFSNHWIF--------WV-----------GPFIGGAL 222

Qy     242 LLCLASGLMLPSCLWFRGTGLIW 264
           : :   :: |    |     :|
Db     223 AVLIYDFILAPRSSDFTDRMKVW 245


RESULT 11
Q9GJX6
ID   Q9GJX6      PRELIMINARY;      PRT;   636 AA.
AC   Q9GJX6;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   SODIUM IODIDE SYMPORTER.
OS   Sus scrofa (Pig).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX   NCBI_TaxID=9823;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=THYROID;
RA   Selmi-Ruby S., Rousset B.;
RT   "Cloning of the pig sodium iodide symporter: Multiple isoforms
RT   generated by alternative splicing.";
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ276292; CAC17816.1; -.
DR   InterPro; IPR001734; Na_solut_symport.
DR   Pfam; PF00474; SSF; 1.
DR   PROSITE; PS00456; NA_SOLUT_SYMP_1; UNKNOWN_1.
DR   PROSITE; PS50283; NA_SOLUT_SYMP_3; 1.
SQ   SEQUENCE   636 AA;  67165 MW;  F80BBFD05152668F CRC64;

  Query Match             3.3%;  Score 122;  DB 6;  Length 636;
  Best Local Similarity   22.4%;  Pred. No. 1.1;
  Matches   71;  Conservative   43;  Mismatches  103;  Indels  100;  Gaps   16;

Qy      15 GCVSWSLYHDLGPMI--------YYFPLQTL----ELTGLEGFSIAFLSPIFLTITPFWK 62
           | | ::|| |  |::         | ||  |    :| |: |        :||
Db     300 GIVMFALYVDCDPLLAGHISAPDQYMPLLVLDIFEDLPGVPG--------LFLACAYSGT 351

Qy      63 LVNKKWMLTLLRIITIGSIASFQAPN-AKLRLMVLALGV-----SSSLIVQAV-TWWSGS 115
           |      :  :  :|:  :   : || |  ||:::: |:     |: | | |: :   |
Db     352 LSTASTSINAMAAVTVEDLIKPRLPNLAPRRLVIISKGLSLIYGSACLTVAALSSLLGGG 411

Qy     116 HLQRYLRIWGFI----LGQIVL------------------VVLRIWY---TSLNPIWSY 149
           ||    : | |    ||  ||                    : | :|     || |  :
Db     412 VLQGSFTVMGVISGPLLGAFVLGMFLPSCNTSGVLSGLAAGLALSLWVAVGASLYPPSAQ 471

Qy     150 QMSNKVILTLSAIATLDRIGTDGDCSKPEEKKTG-----------EVATGMASRPNWLLA 198
           |    :|  ||           |: |    :|            |: : : |   :
Db     472 SMG---VLPSSA----------AGCALPTANASGLQDPVLAVNASSTASSLETDPEQPIL 518

Qy     199 GAAFGSLVFLTHWVFGEVSLVSRWAVSGHPHPGPDPNPFGGAVLLCLASGLMLPSCLWFR 258
           |:| :: :| :   | :| :                ||:: || :|    | |
Db     519 AASFYAISYLYYGALGTLSTI----------------LCGALISCL-TGPTKRSAL--- 557

Qy     259 GTGLIWW--VTGTASAA 273
           | ||:||    ||| |
Db     558 GPGLLWWDLTRQTASVA 574


RESULT 12
Q9GJX7
ID   Q9GJX7      PRELIMINARY;      PRT;   643 AA.
AC   Q9GJX7;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   SODIUM IODIDE SYMPORTER.
OS   Sus scrofa (Pig).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX   NCBI_TaxID=9823;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=THYROID;
RA   Selmi-Ruby S., Rousset B.;
RT   "Cloning of porcine sodium iodide symporter.";
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ276242; CAC17810.1; -.
DR   InterPro; IPR001734; Na_solut_symport.
DR   Pfam; PF00474; SSF; 1.
DR   PROSITE; PS00456; NA_SOLUT_SYMP_1; UNKNOWN_1.
DR   PROSITE; PS50283; NA_SOLUT_SYMP_3; 1.
SQ   SEQUENCE   643 AA;  68127 MW;  29B111986C8E8443 CRC64;

  Query Match             3.3%;  Score 122;  DB 6;  Length 643;
  Best Local Similarity   22.4%;  Pred. No. 1.1;
  Matches   71;  Conservative   43;  Mismatches  103;  Indels  100;  Gaps   16;

Qy      15 GCVSWSLYHDLGPMI--------YYFPLQTL----ELTGLEGFSIAFLSPIFLTITPFWK 62
           | | ::|| |  |::         | ||  |    :| |: |        :||
Db     300 GIVMFALYVDCDPLLAGHISAPDQYMPLLVLDIFEDLPGVPG--------LFLACAYSGT 351

Qy      63 LVNKKWMLTLLRIITIGSIASFQAPN-AKLRLMVLALGV-----SSSLIVQAV-TWWSGS 115
           |      :  :  :|:  :   : || |  ||:::: |:     |: | | |: :   |
Db     352 LSTASTSINAMAAVTVEDLIKPRLPNLAPRRLVIISKGLSLIYGSACLTVAALSSLLGGG 411

Qy     116 HLQRYLRIWGFI----LGQIVL------------------VVLRIWY---TSLNPIWSY 149
           ||    : | |    ||  ||                    : | :|     || |  :
Db     412 VLQGSFTVMGVISGPLLGAFVLGMFLPSCNTSGVLSGLAAGLALSLWVAVGASLYPPSAQ 471

Qy     150 QMSNKVILTLSAIATLDRIGTDGDCSKPEEKKTG-----------EVATGMASRPNWLLA 198
           |    :|  ||           |: |    :|            |: : : |   :
Db     472 SMG---VLPSSA----------AGCALPTANASGLQDPVLAVNASSTASSLETDPEQPIL 518

Qy     199 GAAFGSLVFLTHWVFGEVSLVSRWAVSGHPHPGPDPNPFGGAVLLCLASGLMLPSCLWFR 258
           |:| :: :| :   | :| :                ||:: || :|    | |
Db     519 AASFYAISYLYYGALGTLSTI----------------LCGALISCL-TGPTKRSAL--- 557
```

```
Qy     259 GTGLIWW--VTGTASAA 273
           | ||:||     ||| |
Db     558 GPGLLWWDLTRQTASVA 574


RESULT 13
Q9UYP5
ID   Q9UYP5      PRELIMINARY;      PRT;   766 AA.
AC   Q9UYP5;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   TRANSMEMBRANE OLIGOSACCHARYL TRANSFERASE, PUTATIVE.
GN   PAB0974.
OS   Pyrococcus abyssi.
OC   Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.
OX   NCBI_TaxID=29292;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ORSAY;
RA   Heilig R.;
RT   "Pyrococcus abyssi genome sequence: insights into archaeal chromosome
RT   structure and evolution.";
RL   Submitted (JUL-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ248287; CAB50367.1; -.
DR   InterPro; IPR003674; OTase_STT3.
DR   Pfam; PF02516; STT3; 1.
KW   Complete proteome.
SQ   SEQUENCE   766 AA;  87894 MW;  C7A7E52701819CE4 CRC64;

  Query Match             3.3%;  Score 122;  DB 17;  Length 766;
  Best Local Similarity   23.5%;  Pred. No. 1.4;
  Matches   91;  Conservative   53;  Mismatches  153;  Indels   90;  Gaps   19;

Qy     286 SGCVFAIFTASMWPQTLGHLINSGTNPGKTMTIAMIFYLLEIFFCAWCTAFKFVPGGVYA 345
           :|  |  || :  |   |  | |    |  || | |:  |::|  :  | ||  |
Db      51 AGKWFNFFTIANGP--WGFQIKSFHPLGLWMTPAYIYRFLKVFGISVQTVFKITP----- 103

Qy     346 RERSDVLLGTMMLI---IGLNMLFGPKKNLDLLLQTKNSSKVLFRKSEKYMK----LFLW 398
             |: ||: ::   | |  |:| :|     :    |  :||    | :    :  |
Db     104 -----VIFGTLTIVFFYISLLKLYGKEKAFFASIFLALSFGHIFRSMANYYRGDNYMLFW 158

Qy     399 LLVGVGLLGLGLRHKAYERKLGKVAPT--KEVSAAIWP-----FRFGYDNE-----GWSS 446
           :  :   :   || | | | :  : ||    :|:  |      | |   |      |
Db     159 YSLALAGIAYALRTKGYRRLVFYLVPTLASGISSVFWQAYYPLFVFLSLNGVFLAIGSFL 218

Qy     447 LERSAHLLNETGADFITILESDASKPYMGNNDLTMWLGEKLGFYTDFGPSTRYHTWGIMA 506
           |::  : |:     || || | |    : |     :|||| |:              |::
Db     219 LDKKRNFLD----SFIIIL-STAFGAIIAN-----YLGEKFGY-------------GMLG 255

Qy     507 LSRYPIVKSEHHLLPSPEGEIAPAITLTVNISGKLVD---FVVTHFGNHEDDLDRKLQAI 563
           :|  ||                  | : : ||: |   |:  |:      |   :  :
Db     256 YNRQHIVSK-----------------LGIKL-GKIRDAYLFIHLHY-LVPISLGLLIILL 296

Qy     564 AVSKLLKSSSNQ---VIFLGYIT------SAPGSRDYLQLTEHGNVKDIDSTD----HDR 610
           :|: :||   :   || || ::       |  |  | : :      |  |     ||
Db     297 FLSRFVKSKRAKVGIVIGLGTVSILIIILLKFPALRGLLGIFDMFKSTPIMETRPTNFHDL 356

Qy     611 WCEYIMYRGLIRLGYARISHAELSDSE 637
           |  : :   |: | : |  | |   :|
Db     357 WKAFSISIFLLPLFFLRF-HPEKVKTE 382


RESULT 14
Q36149
ID   Q36149      PRELIMINARY;      PRT;   591 AA.
AC   Q36149;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   NADH-UBIQUINONE OXIDOREDUCTASE SUBUNIT 5.
GN   ND5.
OS   Trachemys scripta (Red-eared slider turtle) (Pseudemys scripta).
OG   Mitochondrion.
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Testudines; Cryptodira; Testudinoidea; Emydidae; Trachemys.
OX   NCBI_TaxID=34903;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=HEART;
RX   MEDLINE=97054595; PubMed=8898892;
RA   Cai Q., Storey K.B.;
RT   "Anoxia-induced gene expression in turtle heart. Upregulation of
RT   mitochondrial genes for NADH-ubiquinone oxidoreductase subunit 5 and
RT   cytochrome c oxidase subunit 1.";
RL   Eur. J. Biochem. 241:83-92(1996).
CC   -!- CATALYTIC ACTIVITY: NADH + UBIQUINONE = NAD(+) + UBIQUINOL.
DR   EMBL; U49048; AAC60055.1; -.
DR   InterPro; IPR003916; NADHub_oxdrdctse5.
DR   InterPro; IPR001750; Oxidored_q1.
DR   InterPro; IPR001516; Oxidored_q1_N.
DR   InterPro; IPR001412; tRNA-synt_I.
DR   Pfam; PF00361; oxidored_q1; 1.
DR   Pfam; PF00662; oxidored_q1_N; 1.
DR   PRINTS; PR01434; NADHDHGNASE5.
DR   PROSITE; PS00178; AA_TRNA_LIGASE_I; UNKNOWN_1.
KW   Mitochondrion; NAD; Oxidoreductase; Ubiquinone.
SQ   SEQUENCE   591 AA;  66062 MW;  1C577C2B77EF901A CRC64;

  Query Match             3.3%;  Score 121.5;  DB 8;  Length 591;
  Best Local Similarity   20.0%;  Pred. No. 1.1;
  Matches  145;  Conservative   86;  Mismatches  226;  Indels  269;  Gaps   36;

Qy       1 MTSLWREILLESLLGCVSWSLYHDLGPMIY--YFPLQTL---ELTGLEGFSI-------- 47
           |: : : :|:      |  : :  | |:::   || |||    | |   |||
Db       5 MSLMSKALLILKTESTVKMTFFISLFPLLHSMLFPEQTLCDVTLLGTSLFSIKLNFIHDQ 64

Qy      48 --AFLSPIFLTITPFWKLVN-KKWMLTLLRIITIGSIASFQAPNAK-------LRLMVLA 97
             |   || | :|  | ::   :| :                |:          : || :
Db      65 YSAIFMPIALYVT--WNILQYSRWYM-------------MDNPHIDKFFKYLLIFLMAMM 109

Qy      98 LGVSSSLIVQAVTWWSGSHLQRYLRIWGFILGQIVLVVLRIWYTSLNPIWSYQMSNKVIL 157
           : ::|: : |    | ||                                    || ::|
Db     110 ILITSNNLFQFFIGWEGS-----------------------------------SNHILL 133

Qy     158 T-----------LSAIATL--DRIGTDGDCSKPEEKKTGEVATGMASRPNWLLAGAAFGS 204
           |            ||:  :  :|||  |            :  ||     ||::     |
Db     134 THWMVTRREDTLASALMAIIYNRIGDLG-----------LYIGMV----WLISKT--NS 175

Qy     205 LVFLTHWVFGEVSLVSRWAVSGHPHPGPDPNPFGGAVLLCLASGLMLPSCLWFRGTGLIW 264
           | |                  |    ||   :::: :|                ||
Db     176 LNF--------------------QPPSYTNP---SLVMLIA--------------LIL 196

Qy     265 WVTGTASAAGLLYLHTW-AAAVSGCVFAIFTASMWPQTLGHLINSGTNPGKTMTIAMIFY 323
           || ::  |   || |  ||: |           |  :  |::|      || :| ||
Db     197 AATGKSAQFG---LHPWLPAAMEG----------PTPVSALLHS-----STMVVAGIFL 237

Qy     324 LLEIF-----------FCAWCTAFKFVPGGVYARERSDV----------LLGTMMLIIGL 362
           |  | |           |    |       : |  ::|:           || ||: |||
Db     238 LIRIHPLLASNNTALSTCLCLGAITTFFASLSALSQNDIKKIIAYSTTSQLGLMMVTIGL 297

Qy     363 NMLFGPKKNLDLLLQTKNSSKVLFRKSEKYMKLFLWLLVGVGLLGLGLRHKAYERKLGKV 422
           |          ||        |     : |  |::  |  :  ||  ::   ||:| :
Db     298 N-------QPDLAF--------LHICMHAFFKAMLFMCAGFIIHSLG--NEQDIRKMGGL 340

Qy     423 -APTKEVSAAI---------WPFRFGYDNEGWSSLERSAHLLNETGADFITILESDASKP 472
            |    || :          ||  || ::       |   ||   |  :|:: :  :
Db     341 HKPLPITSACLTIGNMALMGMPFLTGYYSKDAIIETMSTSYLNSC-ALLLTLMATSFTAV 399

Qy     473 YMGNNDLTMWLGEKL---------GFYTDFGPSTRYHTWGIMALSRYPIVKSEHHLLPSP 523
           |     | :  |  |            |    | ||  |  |||     ::     :  :|
Db     400 YTLRIMLLVQTGHPLQNSTYLLHENIPTIMNPITRLATGSIMA---GWLITKNIQMFNTP 456

Qy     524 EGEIAPAI---TLTVNISGKLVDFVVTHFGNHEDDLDRKLQAIAVSKLLK----SSSNQV 576
           :   |    ||| : | ||                |:  :|  | ||        :
Db     457 TMTMPMWIKIAALTVTVIGLLVGL-------------KVVCMATKKPLKPYKTCDPTSM 502
```

```
Qy     577 IFLGYITSAPG-SRDYLQLTEHGNVKDIDSTDHDRWCEY------------------IMY 617
           |  : | |   : ||:: :       |::    | ||                   :|
Db     503 AGLSQLPSPPQYIKKYLKMAQKIATHLADTS----WFEYLGPKVTAKWQKTPMRFTSLMQ 558

Qy     618 RGLIRL 623
           :|||::
Db     559 KGLIKI 564


RESULT 15
Q9B6Y4
ID   Q9B6Y4      PRELIMINARY;      PRT;   605 AA.
AC   Q9B6Y4;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   NADH DEHYDROGENASE SUBUNIT 5.
OS   Casuarius bennetti (Dwarf cassowary).
OG   Mitochondrion.
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Palaeognathae; Casuariiformes; Casuariidae;
OC   Casuarius.
OX   NCBI_TaxID=30463;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=MUSCLE;
RX   MEDLINE=21085666; PubMed=11217857;
RA   Cooper A., Lalueza-Fox C., Anderson S., Rambaut A., Austin J.,
RA   Ward R.;
RT   "Complete mitochondrial genome sequences of two extinct moas clarify
RT   ratite evolution.";
RL   Nature 409:704-707(2001).
CC   -!- CATALYTIC ACTIVITY: NADH + UBIQUINONE = NAD(+) + UBIQUINOL.
DR   EMBL; AY016011; AAK08552.1; -.
DR   InterPro; IPR003916; NADHub_oxdrdctse5.
DR   InterPro; IPR001750; Oxidored_q1.
DR   InterPro; IPR001516; Oxidored_q1_N.
DR   Pfam; PF00361; oxidored_q1; 1.
DR   Pfam; PF00662; oxidored_q1_N; 1.
DR   PRINTS; PR01434; NADHDHGNASE5.
KW   Mitochondrion; NAD; Oxidoreductase; Ubiquinone.
SQ   SEQUENCE   605 AA;  66408 MW;  41B286D9EC828A26 CRC64;

  Query Match             3.3%;  Score 121;  DB 8;  Length 605;
  Best Local Similarity   19.6%;  Pred. No. 1.2;
  Matches  139;  Conservative   87;  Mismatches  230;  Indels  254;  Gaps   34;

Qy      27 PMIYYFPLQTLELTGLEGFSI-------AFLSPIFLTITPFWKLVNKKWMLTLLRIITIG 79
           ||     ::|   | |   ||          :|  :   || :|:
Db      35 PMSITSTVKTAFFTSLIPMSIFIYSGTETIVSQWYWNFTPNFKIP--------------- 79

Qy      80 SIASFQAPNAKLRLMVLALGVSSSLIVQAVTWWSGS--HLQR-YLRIWGFILGQIVLVV- 135
             ||:     :    :|| |: | |:|  ||:  |  |: : :: :  |::  : | :
Db      80 --LSFKMDQYSMLFFPIALFVTWS-ILQFATWYMASEPHITKFFIYLLTFLIAMLTLTIA 136

Qy     136 -----LRIWYTSLNPI------WSYQMSNKVILTLSAIATLDRIGTDGDCSKPEEKKTGE 184
                | :  :  :      | :  :      | |:   :|||  |
Db     137 NNMFMLFIGWEGVGIMSFLLIGWWHGRAEANTAALQAV-IYNRIGDVG------------ 183

Qy     185 VATGMASRPNWLLAGAAFGSLVFLTHWVFGEVSLVSRWAVSGHPHPGPDPNPFGGAVLLC 244
                           |:    |:    | :: | :    :    |
Db     184 -------------------LILSMAWL---ASTINTWEIQQASYETQMP---------- 210

Qy     245 LASGLMLPSCLWFRGTGLIWWVTGTASAAGLLYLHTW-AAAVSGCVFAIFTASMWPQTLG 303
             |||        |||   || ::  |   || |  ||: |           |  :
Db     211 -----MLPL------LGLILAATGKSAQFG---LHPWLPAAMEG----------PTPVS 245

Qy     304 HLINSGTNPGKTMTIAMIFYLLE----------IFFCAWCTAFKFVPGGVYARERSDV- 351
           |::|      || :| || |:           :  |    |   :     |  ::|:
Db     246 ALLHS-----STMVVAGIFLLIRTHPLLTNNQTALTTCLCLGALSTLFAATCALTQNDIK 300

Qy     352 ---------LLGTMMLIIGLNMLFGPKKNLDLLLQTKNSSKVLFRKSEKYMKLFLWLLVG 402
                    || ||: ||||:                    |   :  : |  |:|  |
Db     301 KIIAFSTSSQLGLMMVTIGLNL----------PQLAFLHISTHAFFKAMLFLCSG 345

Qy     403 VGLLGLGLRHKAYERKLG---KVAPTKEVSAAI-------WPFRFGYDNEGW--SSLERS 450
             :  ||       ||:|   |: ||       |        ||  |: ::     ||  |
Db     346 SIIHSLGGEQDI--RKMGGLQKLLPTTTACLTIGSLALMGTPFLAGFYSKDLIIESLNNS 403

Qy     451 AHLLNETGADFITILESDASKPYMGNNDLTMWLGEKLGFYTDFGPSTRYHTWGIMALSRY 510
             || | |  :|:| :  :  |     : | |  : ||                   :|
Db     404 --YLN-TWALLLTLLATSFTATY----SIRMTLMVQAGF----------------TRI 438

Qy     511 PIVKSEHHLLPSPEGEIAPAIT--------------------------LTVNISGKLV 542
           | :          :|  | :||:|                        :|: :  |:
Db     439 PPI--------TPINENSPAVTNPITRLAVGSIIAGLLITSNVLPTKTPPMTMPMVTKMA 490

Qy     543 DFVVTHFGNHEDDLDRKLQAIAVSKLLKSS-----SNQVIFLGYITSAPGSRDYLQLTEH 597
             :||  |     :   |: : ::  |  |     ||  | |||        : : |  :
Db     491 AIIVTILG-----IMLALELLNLTHNLTPSKQNTYSNFSITLGYFNPLLHRLNPMMLLNN 545

Qy     598 GNVKDIDSTDHDRWCEYIMYRGLIRLG-YARISHAELSDSEIQMAKFRIP 646
           |          |            || |  | ::    |:| ::  ||   |
Db     546 GQ----KIASH-----------LIDLSWYKKMGPEGLADLQLMAAKTSTP 580


Search completed: October 21, 2002, 22:10:57
Job time : 68 secs
```